

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

Conflicting views of the Greenland Ice Sheet's past
offer clues to future sea levels PAGES 202, 252 & 256

MELTING POINTS

REVIEWS

GET THEM WHEN THEY'RE YOUNG

Recommended books for
the budding scientist

PAGE 194

HISTORY

ST HELENA'S DARK PAST

Ancient DNA sheds new
light on the slave trade

PAGE 184

GEOLOGICAL TIME

DEFINING THE ANTHROPOCENE

Social sciences should
be part of the story

PAGE 192

NATURE.COM/NATURE

8 December 2016 £10

Vol. 540, No. 7632



9 770028 083095

THIS WEEK

EDITORIALS

LITERATURE Children's books can mislead on the Moon's phases **p.170**

WORLD VIEW Fight back against junk science on the web **p.171**



VACCINES Genetic tweaks point towards safer flu jabs **p.172**

Pet projects need a helping hand

Clinical trials with cats and dogs offer great promise for animal and human medicine but risk being stifled by overzealous regulations.

For decades, the usual veterinary response to a pet's unbearable suffering has been the same: the dearly loved animal has been quietly and humanely put to sleep. Yet a new trend has emerged over the past decade or so: in search of hope, or just a few extra months of life, owners have been willing to enrol their pets in experimental trials of new therapies. Science and medicine recognize this, and see a splendid opportunity for both pets and people. Rules must now be adjusted to exploit this potential.

Clinical trials of drugs are increasingly being carried out on pets, particularly dogs and cats. Such trials are analogous to those conducted in people, and yield reliable data that can lead to swifter approval and marketing of new veterinary products. The results can also support the much tougher procedures to approve new treatments for related conditions in people.

It should be a win-win situation. When little Fritz develops a nasty lump on her neck, or lovable Tom-tom starts walking with a painful limp, a vet can, after diagnosis, offer to recruit them into any relevant clinical trial — with the possibility of a better-than-standard treatment. Veterinary surgeons say that nearly all pet owners give eager informed consent to participate, either in the hope of exploiting that possibility or because, as serial pet owners, they hope that the research will help their next animal.

A forerunner of this trend, and a continuing gold standard, is the US National Cancer Institute's Comparative Oncology Trial Consortium, which has been running for 12 years and recruits pet dogs into specific cancer trials. A dozen trials have been completed and some have supported pharmaceutical-company decisions to drop or pursue candidate drugs for human use. In the past few years, ambitious veterinary institutions around the world have started their own pet trials for conditions from cancer to arthritis and diabetes — and their focus is on both veterinary and human therapies.

Veterinary surgeons are happy because the trials help to speed approval for treatments for their furry patients. Regulators of human medicines are also enthusiastic. They welcome relevant pet clinical-trial data as part of a drug-developer's evidence that a medicine is safe and effective. Pets also offer some very specific advantages. Most tests involve laboratory animals especially bred or modified to represent key aspects of a disease — but pet animals that actually have the condition are the real McCoy. They are genetically diverse, they develop the disease spontaneously and they share the human environment. So pet trials much more closely reflect the real-life situation for people.

Aside from publicly funded trials in some countries, pharmaceutical companies regularly approach veterinary schools for direct collaboration in their drug-discovery programmes. It's a problem, then, that a regulatory gap threatens this work. Unlike the rules for human clinical trials, regulations for veterinary trials are unclear and confused, especially in the European Union.

There are no international guidelines, so some EU countries,

including Austria and the United Kingdom, have chosen to classify pet clinical trials as animal experimentation. As such, they fall under the (rightly strict) 2010 EU directive on the protection of animals used for scientific purposes.

This adds complications that tie the authorities in knots, delaying approval of medicines. For example, the directive is designed for laboratory animals and sets out precise rules for their housing and routine care, something that by definition cannot be controlled for privately owned pets. Moreover, veterinary surgeons in Austria, a country with highly sensitive attitudes to animal rights, say that the definition of the work as an experiment rather than a trial prompts many pet owners to choose euthanasia.

Trial runners in other countries, including Germany and the United States, have more flexibility. One answer to the dilemma might be for the stricter EU countries to reconsider their rigid positions, and loosen their rules for pet trials.

Another is for scientists and veterinary surgeons to lobby for an amendment to the EU directive itself, so that it explicitly excludes pet clinical trials — in the same way that it already excludes trials required for marketing authorization. The legislation is currently under review, so now would be the right time to make the change.

Pet work attracts cute headlines, but it is no curiosity. The research is valuable and deserves support — from both researchers and policymakers. ■

“Complications tie the authorities in knots, delaying approval of medicines.”

Ebb and flow

Humans cause most changes in Earth's surface water, so its governance must become a priority.

“Human nature is like water,” the US poet Wallace Stevens wrote. “It takes the shape of its container.” As the political events of 2016 continue to raise questions about that shape and how it is changing, it might seem difficult to redirect attention to water. So delegates gathering in Rabat, Morocco, next month for the eighth meeting of the Water Governance Initiative, run by the Organisation for Economic Co-operation and Development (OECD), deserve acknowledgement at the very least. At the meeting, which falls in the week before the next US president is sworn into office, those hardy souls are set to discuss “raising the profile of water governance in the Global Agenda”.

Yet the importance of water governance, if not its political profile,

increases steadily year on year. And the politics of water — for that is what its governance involves — could yet hold a lesson for more-solid human affairs. Even the definition of water governance is political, and a hard-won human compromise, so it can seem a bit, well, fluid: the social, economic and political systems that control decision-making on water-resource development and management.

Its goal is surely one that all can agree on: to make sure that people do not have too much water, nor too little, and that it's not too polluted. But the ways of achieving that across the globe mirror the governance of water turned into snowflakes: no two circumstances are identical.

A 2013 study, for instance, reported on the introduction of sustainable practices to maintain water flow in the local environment (rather than piping it away for human use) at river basins in China and Australia (R. Q. Grafton *et al. Nature Clim. Change* 3, 315–321; 2013). Whereas changes to China's Yellow River were imposed by the central Communist government, improvements in the Murray–Darling River basin were the product of a market-based system that encouraged the trading (and non-use) of extraction permits.

Those are the (limited) success stories. But the harsher reality of politics frequently pours into water governance. And as a microcosm of the wider world, different attempts to manage water resources often serve only to make other options seem more attractive. Perhaps more than for any other natural resource, the various groups in society can see something different reflected back when they look at water. It's a source of life, hygiene, crops, leisure, industry, livelihood — or just a pretty view — and those interpretations often collide. It's easy to see why scholars of water governance typically prefer to focus on the actors rather than on the actions that are needed.

That's another reason why the Morocco conference next month should be recognized. It is the latest in a long process that aims to guide policies of water governance. Run under the umbrella of the OECD, the Water Governance Initiative works to set principles and share good practice. It aims, for example, to increase the number of river basins

that are assessed and given management plans, and to encourage countries to identify and crack down on corruption in the water business.

It is making progress. Last year, the initiative agreed on a dozen principles to guide water governance, and is now consulting on suitable indicators that could be used to measure progress. Scientists can do their bit here: one of the key principles is the need for policy-relevant and timely data and information on water use and resources.

“Many lakes and rivers have disappeared in our own lifetimes.”

Indeed, some such information appears in a *Nature* research paper published online this week (J.-F. Pekel *et al. Nature* <http://dx.doi.org/10.1038/nature20584>; 2016) that offers the most comprehensive picture yet of the planet's water resource. The study collects and collates more than 3 million satellite images of

Earth's surface taken over the past three decades, and shows how surface water — rivers, lakes and wetlands — has ebbed and flowed. If there is, or was until recently, a pool of open water at least 30 square metres near you, then it's probably included in this map. The pictures reveal some big numbers: nearly 4.5 million square kilometres of the planet's land surface has been under water at some point since 1984 (still just 3%). More than half of this is in the global north, above a latitude of 44° N.

Although we think of lakes and rivers as features of the landscape, many have disappeared in our own lifetimes. More than 90,000 km² of water bodies thought to be permanent have disappeared — including giant chunks of the Aral Sea — and a further 72,000 km² is now classed as only seasonally flooded. But overall, there has been more flow onto the land than away from it: almost 213,000 km² of land that was dry in 1984 is now covered in water for some periods, often inside new reservoirs.

Climate change has a role in these shifts, but the biggest cause of water movement is direct human activity. Some regions have too much, others too little. Water governance deserves its place on the global agenda before it's too late. ■

Word power

How the books children read can trigger a lifelong fascination with science.

Even *Nature* editors were young once. And in a Books & Arts special this week, *Nature* readers can get a little insight into what makes those editors tick — or, at least, what helped to turn them on to science (see page 194). Forget virtual-reality goggles and computer-simulated, movement-sensitive interactions. These were the olden days, and stories of the world reached our young editors from inside the covers of books.

Ask someone what reading material inspired them as a child, and few will be honest enough to say that it was comics such as the *Beano*. So we have to trust our editors when they say they spent their formative years curled up with childhood investigations of chemistry, physics, mathematics and, chiefly, the natural world. But reading their recollections — and, indeed, how some read more-modern works with their own children — it's easy to see why they did so.

The pages they describe tell science as opportunity and discovery, learning without instruction, and of fascination and imagination. More, it is science as embedded in society and the world, science as a relevant, integral, natural and core ingredient for a curious and active mind. With a foundation like that, it's easy to see why *Nature* editors and readers sometimes struggle to understand how anyone would choose to see the world in any other way.

To analyse the content of children's science books can be like pulling the beard of Father Christmas to see if it's real. Some of the magic is

lost in the process. But there is a serious — and an educational — side to children's books. Their influence is great, and as such their style and content have been scrutinized over everything from their depictions of violence and gender roles, to people's attitudes to the environment and recycling. If books leave such a lasting impression on people, then should scientists and researchers do more to make sure that those read to and by children are accurate? Where, for example, does fiction tip into fantasy — and should young readers be made aware of the difference?

These concerns can seem overblown and ripe for ridicule. Does the anthropomorphism of cats and dogs in stories make children believe that their pets can talk? (Probably not.) But can representations of humans as superior and somehow distinct from other animal species fuel misconceptions about our origins? (Perhaps.)

For an example of the power of children's literature to mislead, simply look up at the sky. The phases of the Moon is a tricky concept to grasp at first, and studies show that even university science graduates mistakenly attribute them to Earth's shadow. (In fact, as you know, the Moon's phases reflect how much of its sunlit half is visible from Earth.) Educationists put much of the blame for this on the way in which the Moon is shown in children's books, including the much-loved Eric Carle classic *Papa, Please Get the Moon for Me*.

The literary lunar laxness goes even deeper. If children and many adults are sometimes surprised to see the Moon in the daytime, then some of the blame for that can be placed on kids' books, too. (And so, but less commonly, can the expectation that a cow will leap over it.)

The Moon aside, children's books are rightly viewed as a greatly valued resource of science. Yes, even those of Eric Carle. Kids may not grasp celestial mechanics, but most know the life cycle of the butterfly. That's assuming, of course, they have read Carle's *The Very Hungry Caterpillar*. ■



Take the time and effort to correct misinformation

Scientists should challenge online falsehoods and inaccuracies — and harness the collective power of the Internet to fight back, argues Phil Williamson.

With the election of Donald Trump, his appointment of advisers who are on record as dismissing scientific evidence, and the emboldening of deniers on everything from climate change to vaccinations, the amount of nonsense written about science on the Internet (and elsewhere) seems set to rise. So what are we, as scientists, to do?

Most researchers who have tried to engage online with ill-informed journalists or pseudoscientists will be familiar with Brandolini's law (also known as the Bullshit Asymmetry Principle): the amount of energy needed to refute bullshit is an order of magnitude bigger than that needed to produce it. Is it really worth taking the time and effort to challenge, correct and clarify articles that claim to be about science but in most cases seem to represent a political ideology?

I think it is. Challenging falsehoods and misrepresentation may not seem to have any immediate effect, but someone, somewhere, will hear or read our response. The target is not the peddler of nonsense, but those readers who have an open mind on scientific problems. A lie may be able to travel around the world before the truth has its shoes on, but an unchallenged untruth will never stop.

Earlier this year, I had a run-in with Breitbart News — the libertarian website made infamous by the appointment of its former senior executive Stephen Bannon as Trump's chief strategist. It followed an article in *The Spectator* that criticized research on ocean acidification and contained several inaccuracies, written by James Delingpole, who also edits Breitbart London. To give an idea of the standard of discussion, Delingpole argued that there has been no long-term reduction in ocean pH levels and that future climate change would cause the release of carbon dioxide from the ocean. Acidification is therefore a non-problem invented by 'climate alarmists' because there is insufficient evidence for global warming.

I coordinated the UK research programme on ocean acidification and have been involved in national and international evidence assessments. There are genuine scientific uncertainties, but those were not the issues raised by *The Spectator*. When I complained to the magazine, no acknowledgement was received. I then published a rebuttal in *The Marine Biologist*, which prompted Delingpole to write on the Breitbart site that my work should be squashed like a slug.

I referred my complaint about the *Spectator* article to the UK Independent Press Standards Organisation (IPSO) and I expect its verdict soon. Issues of dispute include whether, as the article states, "marine life has nothing whatsoever to fear from ocean acidification" and whether "as looks increasingly to be the case, that the man-made global warming theory is a busted flush".

There is overwhelming scientific evidence to contradict those statements — strongly based on experiments, field observations and theory.

So it would be a defeat of rationality and science if my complaint to IPSO is dismissed as a matter of opinion, rather than fact. Nevertheless, results from around 100 researchers involved in the UK national programme on ocean acidification (comprising about 250 peer-reviewed papers) were disregarded in the *Spectator* article as providing no more new knowledge than could be obtained from "a few hours' basic research".

Brandolini's law certainly applies in this case. Demonstration of (incontrovertible) factual inaccuracy through the IPSO complaints procedure takes time and effort. But I think it is worth it — the scientific process doesn't stop when results are published in a peer-reviewed journal. Wider communication is also involved, and that includes ensuring not only that information (including uncertainties) is understood, but also that misinformation and errors are corrected where necessary.

Most scientists who see their work misrepresented don't have access to a formal complaints procedure. Online journalism seems to be subject to few, if any, rules. Only UK websites directly linked to IPSO-regulated media are subject to its standards. And even for those that are regulated, IPSO's previous incarnation, the Press Complaints Commission, was reluctant to become involved: "In the realm of blogging... there is likely to be strong and fervent disagreement, with writers making use of emotive terms and strident rhetoric. This is a necessary consequence of free speech."

Yet the rising tide of populism threatens the future of evidence-based governance. Social media and websites, lacking quality control, are replacing newspapers as the main information

sources for the public and many politicians, even at the highest level. Not much can be done about this nonsense online, but 'not much' is still something. Discussion threads provide some opportunity for challenge and informed comment, but are not for the faint-hearted. Another approach is fact-checking websites such as Climate Feedback. Scientists should support such efforts, even though they don't have the resources to scrutinize every dubious claim.

For greatest effect, I suggest that we harness the collective power and reach of the Internet to improve its quality. The global scientific community could learn from websites such as travel-review site TripAdvisor, Rotten Tomatoes (which summarizes film and play reviews) and alexa.com (which quantifies website popularity), and set up its own, moderated, rating system for websites that claim to report on science. We could call it the Scientific Honesty and Integrity Tracker, and give online nonsense the SHAIT rating it deserves. ■

Phil Williamson is an associate fellow at the University of East Anglia in Norwich, UK.

e-mail: p.williamson@uea.ac.uk

THE RISING TIDE OF
POPULISM
THREATENS
THE FUTURE OF
EVIDENCE-BASED
GOVERNANCE.

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

CONDENSED-MATTER PHYSICS

Frozen bismuth superconducts

Bismuth crystals can act as superconductors — but only at temperatures approaching absolute zero.

Researchers have long been searching for new superconducting materials, which have the ability to maintain an electric current with no power source. A team led by Srinivasan Ramakrishnan at the Tata Institute of Fundamental Research in Mumbai, India, showed that bismuth crystals become superconducting if they are brought down to 0.5 millikelvin. This extreme cooling is necessary because, unlike other superconductors, bismuth has a very low density of charge carriers, with just 1 electron per 100,000 atoms.

The standard theory of superconductivity can explain the phenomenon in other metals, such as lead and aluminium, but not in bismuth, so further study is needed, the authors say.

Science <http://doi.org/bvbj> (2016)

VACCINES

New way to tame a virus

A live, genetically modified flu virus can infect animals and trigger a strong immune response, but cannot multiply

in its host's cells. Such modified viruses could one day be used to improve on current vaccines (**pictured**).

Vaccines made of live viruses elicit stronger protective immune responses than inactivated vaccines, but, because they can replicate, have the potential to cause disease. To overcome this, Demin Zhou and his colleagues at Peking University in Beijing genetically altered the influenza A virus so that it

could be produced efficiently by special transgenic cells, but could not replicate in normal cells or in infected animals. When compared with a commercially available inactivated flu vaccine, the modified virus stimulated stronger immune reactions in mice, ferrets and guinea pigs. Mice given the new vaccine and then infected with the unmodified flu virus survived, whereas all unvaccinated mice died.

Science 354, 1170–1173 (2016)

MICROBIOLOGY

Gut bacteria linked to Parkinson's

Bacteria living in the gut may contribute to movement problems seen in disorders such as Parkinson's disease.

Timothy Sampson and Sarkis Mazmanian at the California Institute of Technology in Pasadena and their team generated mice that lacked their own bacteria and had been genetically engineered so



MAURICIO DE PAIVA

PALAEoANTHROPOLOGY

Early burials had mutilation rituals

Ancient humans in South America used complex funeral rituals and manipulated the bodies of their dead as early as 10,000 years ago.

André Strauss at the University of Tübingen in Germany and his colleagues analysed bones from 26 human burials, discovered during 15 years of archaeological excavations in a cave in east-central Brazil (pictured). They found that the remains were treated before and during burial in a variety of ways, including defleshing

and tooth removal. Bones dating to roughly 9,500 years ago showed signs of dismemberment and burning before being carefully arranged, suggesting a ceremonial burial. Similar practices were known in the Andes from this period but not in eastern South America.

The authors say these ritualized burials may have been important in maintaining social cohesion in times of stress and conflict.

Antiquity 90, 1454–1473 (2016)



BRIAN SNYDER/REUTERS

SEVEN DAYS

The news in brief

ENERGY

EU clean energy

The European Commission on 30 November proposed a sweeping set of measures to improve overall energy efficiency in the European Union by 30% by 2030. The Clean Energy for All Europeans package includes new standards for energy-friendly building and energy labelling of consumer goods. The commission hopes that the proposed policies and incentives will mobilize up to €177 billion (US\$188 billion) in public and private investment in energy-efficient construction, transport and manufacturing. The package updates a binding 2012 EU directive that commits the bloc to a 20% improvement in energy efficiency by 2020.

POLICY

Gag clause dropped

The UK government has abandoned restrictions on government funding that critics said would have amounted to a 'gagging clause' on academics. Proposals to prevent government money from being used for broadly defined lobbying activities were unveiled in February. But they were put on hold after scientists protested that activities such as giving expert evidence to parliament would be included. On 2 December, the government finalized its grant standards and confirmed that scientific advice and publicizing research results would not be restricted. Science advocates have welcomed the news.

UK fertility step

The United Kingdom is closer to becoming the first country to explicitly permit the birth of children from embryos that contain three

people's DNA. Scientists advising the UK Human Fertilisation and Embryology Authority (HFEA) said on 30 November that mitochondrial-replacement therapy — which aims to erase diseases transmitted by the DNA found in cellular structures called mitochondria — is ready for limited clinical testing. Britain legalized such therapy in 2015, but the HFEA had asked for more research before allowing clinical trials. If it gives the go-ahead when it meets on 15 December, the first procedures, approved on a patient-by-patient basis, could occur early next year.

PUBLISHING

arXiv upgrade

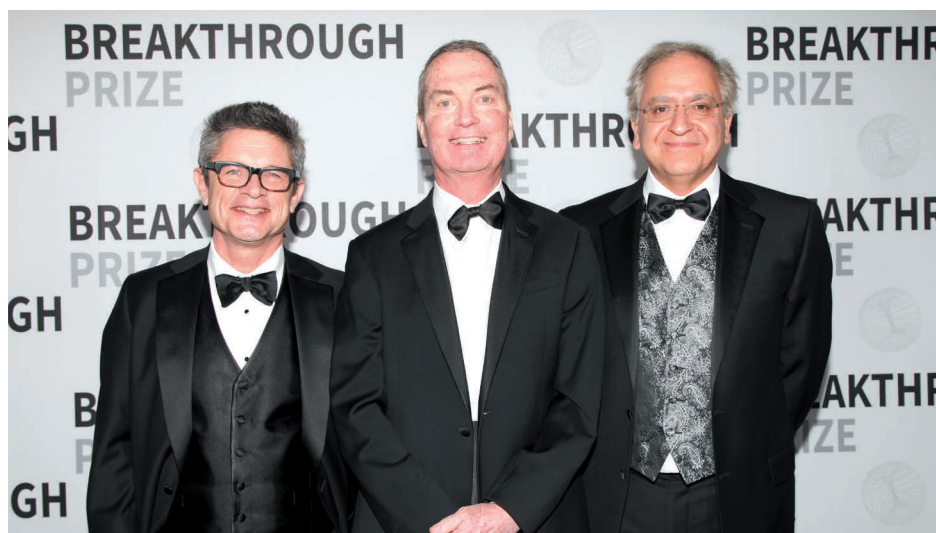
The popular open-access physics repository arXiv is beginning the first phase of a three-year overhaul. With the help of a US\$445,000 grant from the Alfred P. Sloan Foundation in New York City, announced on 29 November, the repository's owners will start planning a 'next-generation arXiv', to be known as arXiv-NG — a more flexible service that can accommodate new features. In June, the site's leaders revealed the results of a massive user survey, which it is using to guide its modernization. Most users

called for no major changes but suggested a few tweaks, such as to the site's search functions.

EVENTS

LIGO goes again

LIGO, the Laser Interferometer Gravitational-wave Observatory has resumed its search for gravitational waves, starting its second science run on 30 November after upgrades to make the system more sensitive. In September 2015, LIGO made the first direct detection of gravitational waves, from a collision between distant black holes. The upgrades should make detections of the phenomenon



KELLY SULLIVAN/GETTY FOR BREAKTHROUGH PRIZE

Black-hole puzzle nets \$3-million prize

The discovery of a physics puzzle known as the black-hole firewall paradox has bagged three string theorists — Andrew Strominger, Joseph Polchinski and Cumrun Vafa (pictured, left to right) — one of the US\$3-million Breakthrough Prizes, the most lucrative awards in science. The annual prizes, together worth \$25 million, honour achievements in fundamental physics, life sciences and mathematics. US molecular biologist Harry Noller won one of five life-sciences prizes, for revealing the centrality of

ribosomal RNA to protein synthesis. The maths award went to US-based Jean Bourgain for work on the geometry of multidimensional spaces and on solving partial differential equations. A special prize was awarded to the 1,015 people working on the LIGO project, which last year made the first detection of gravitational waves. A glitzy awards ceremony took place at NASA's Ames Research Center in Mountain View, California, on 4 December. See go.nature.com/2gxims3 for more.



more frequent. LIGO has twin detectors in Hanford, Washington, and Livingston, Louisiana (**pictured**).

New element names

Four new element names — nihonium (element 113), moscovium (115), tennessine (117) and oganesson (118) — were confirmed by the International Union of Pure and Applied Chemistry (IUPAC) on 28 November. The names, outlined in June by IUPAC, were proposed by the labs credited with discovering the elements — in Russia, the United States and Japan. Yuri Oganessian becomes only the second living person to have an element named after them.

Dear Mr Trump

More than 2,300 scientists, including 22 Nobel laureates, from all 50 US states published an open letter on 30 November to the incoming government

of US president-elect Donald Trump. They urge the administration to create a strong culture of science and recognize that diversity makes science stronger; to ensure that environmental laws retain a strong scientific foundation; and to provide adequate resources for scientists to do their jobs. The scientific community stands ready “to hold accountable any who might seek to undermine” the role of science in policymaking, says the letter.

AWARDS

Mentoring prize

This year's *Nature* Awards for Mentoring in Science have been given to two mentors from the US west coast. Julie Overbaugh, an HIV researcher at the Fred Hutchinson Cancer Research Center in Seattle, Washington, was awarded the lifetime-achievement

prize, and biologist Susan Forsburg at the University of Southern California, Los Angeles, received the prize for mid-career achievement. Each receives US\$10,000. The awards honour outstanding scientific mentors in a different country or region each year. See page 305 for more.

PEOPLE

Prize cash donated

A Kenya-based malaria researcher who won a US\$1-million development prize will donate the money to African science. Kevin Marsh at the University of Oxford, UK, won the inaugural Kuwait-sponsored Al-Sumait Prize for individuals or institutions who advance economic, social and human-resources development in Africa on 22 November. He is a senior adviser to the African Academy of Sciences and led the renowned Kenya Medical Research Institute in Nairobi from 1989 to 2014. Marsh will donate \$400,000 to the academy to create a fund to support young scientists, and the remaining \$600,000 to Oxford's African science programmes.

FUNDING

Europe to Mars

Europe is set to land its first rover on Mars in 2021, after the European Space Agency

COMING UP

12–16 DECEMBER

The American Geophysical Union's Fall Meeting, the largest Earth and space-science meeting in the world, takes place in San Francisco, California. fallmeeting.agu.org

16 DECEMBER

SpaceX's Falcon 9 rocket returns to flight after a failure in September, to launch satellites for communications company Iridium from California. go.nature.com/2gl2tey

secured the €436 million (US\$464 million) it needs for the mission at a meeting in Switzerland on 2 December. European government ministers agreed to provide the cash despite rising costs and the failure of the agency's Mars lander Schiaparelli in October. ExoMars 2020 will be a joint mission with Russia. See page 177 for more.

EU defence fund

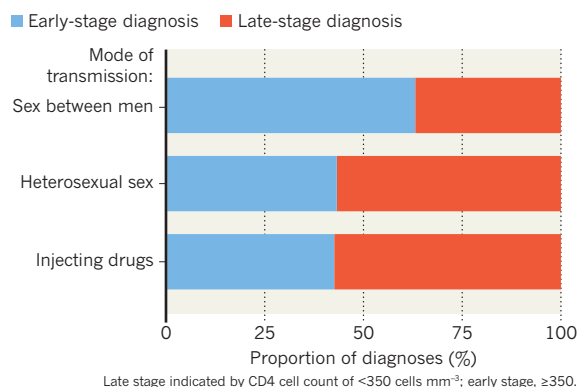
The European Commission plans to launch a fund dedicated to defence research, it announced on 30 November. The research programme will cover technologies such as electronics, encrypted software and robotics and will be run under a broader European Defence Fund. It will have an expected budget of €90 million (US\$97 million) over three years from 2017 and will be separate from the European Union's Horizon 2020 research programme. Classified calls will be open only to European Union member states and Norway. If the scheme is successful, the Commission hopes to increase the programme's budget to around €500 million a year from 2021.

TREND WATCH

Nearly 1 in 7 people with HIV in Europe do not know they have it, estimates a report from the European Centre for Disease Prevention and Control. Delayed diagnoses mean that some may not benefit from treatment and may spread the virus. The data suggest that people contracting HIV through injecting drugs or heterosexual sex are likely to be diagnosed later than those who acquire it through sex between men (see chart). The average time from infection to diagnosis is almost four years.

LATE HIV DIAGNOSES PREVALENT IN EUROPE

Nearly half (47%) of HIV-positive people in the European Union and European Economic Area are diagnosed late. The infection rate in 2015 was 6.3 per 100,000, similar to recent years.



NEWS IN FOCUS

ANCIENT GENOMES DNA analysis delivers Spirit Cave Mummy to US tribe **p.178**

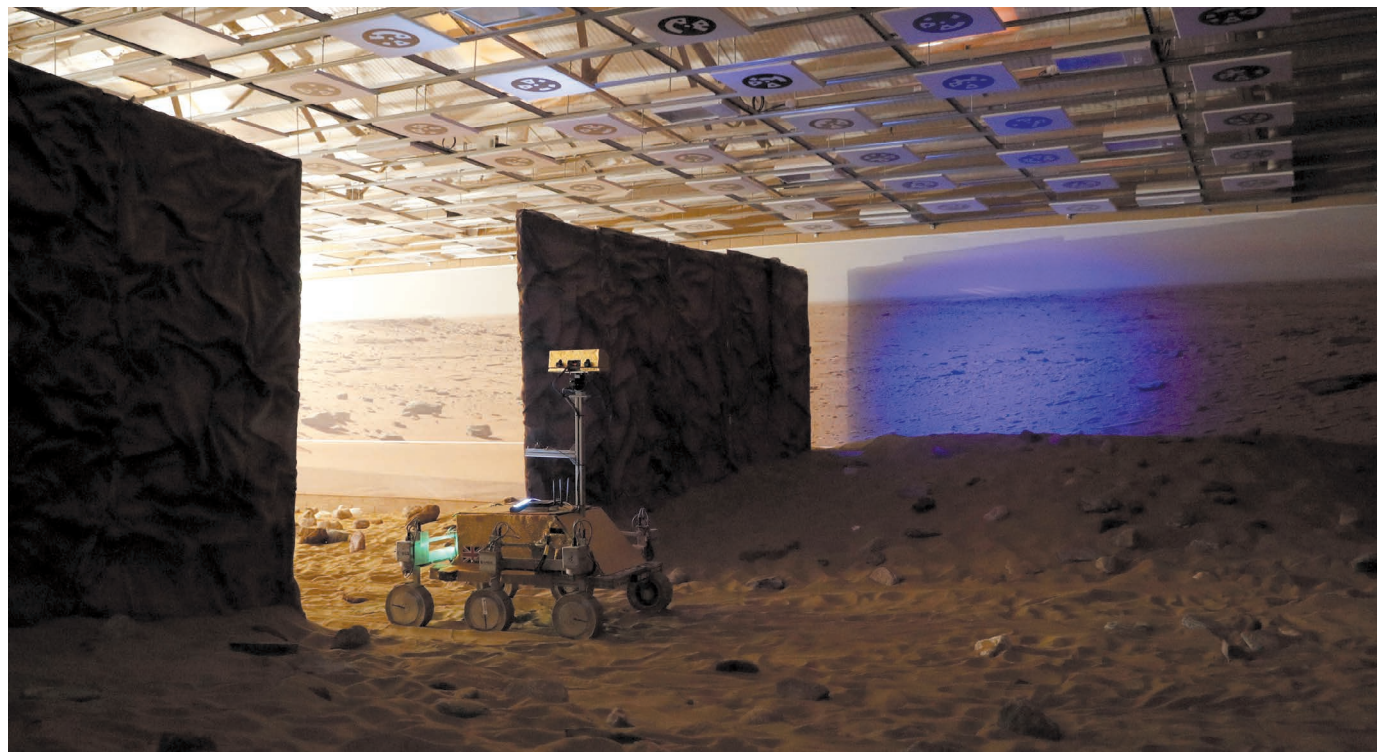
MEXICO Proposal to ban embryo research could stifle science **p.180**

SOCIOLOGY Academics struggle to explain wave of nationalism **p.182**



GENOMICS DNA used to trace the roots of freed slaves **p.184**

ADRIAN DENNIS/AFP/GETTY



A prototype of the ExoMars rover explores a simulated Mars environment.

PLANETARY SCIENCE

Mars rover goes ahead despite crash of test craft

But the European Space Agency's role in a mission to deflect an asteroid moon is over.

BY ELIZABETH GIBNEY

Europe's space agency has secured the cash that it needs to land its first rover on Mars — despite rising costs and the failure of its test Mars lander, Schiaparelli, in October.

“After the many challenging, difficult and rewarding moments of 2016, this is a great relief and a fine result for European space exploration,” says Don McCoy, project manager for the overall ExoMars project, which includes the rover and Schiaparelli, at

the European Space Agency (ESA).

The launch of the Mars rover, a joint project with the Russian space agency Roscosmos, had already been delayed to 2020 from a planned 2018 — a major cause of the cost increase. And the future of the rover mission, called ExoMars 2020, looked to be on shakier ground following the failure of Schiaparelli, which was designed to test Europe's ability to land a craft on the red planet.

But at a meeting of European government ministers in Lucerne, Switzerland, on 1–2 December, ESA member states agreed to

provide an extra €339 million (US\$365 million) for ExoMars 2020, on top of the more than €1 billion already committed to the ExoMars programme. ESA also announced that it will find a further €97 million by moving funds internally. Speaking at a press briefing after the meeting, ESA director-general Jan Wörner said that this would be done “without detriment” to ESA's wider science budget. Together, these sums provide the €436 million needed for ExoMars 2020 to go ahead.

ESA hopes to learn lessons from Schiaparelli. Last week, the agency confirmed ►

► that the crash was caused by errors in the sensor systems that made the craft behave as if it was closer to the Martian surface than it really was. This prompted Schiaparelli to jettison its parachute too early, before falling from a height of about 3.7 kilometres.

“We will have learned much from Schiaparelli that will directly contribute to the second ExoMars mission,” said David Parker, ESA’s director of human spaceflight and robotic exploration at the European Space Research and Technology Centre in Noordwijk, the Netherlands, in a statement.

If successful, the rover will be the first that either Europe or Russia has operated on Mars’s surface. Its major selling point is a drill that can extend to 2 metres, which would allow the robot to dig to a depth where signs of ancient life could have been preserved, protected from the red planet’s harsh radiation.

ASTEROID FAIL

But the meeting also brought bad news for another project. Member states did not commit the €250 million needed to fund a plan for ESA to participate in a mission to deflect the moon of an asteroid.

The Asteroid Impact Mission (AIM) was designed to test a strategy for protecting Earth

from incoming asteroids by subtly changing their course using projectiles. Insiders say that AIM missed its target sum by perhaps a few tens of millions of euros.

“A cool project has been killed because of a lack of vision — even short term — and courage, and this is really sad,” says Patrick Michel, a planetary scientist at the French National Centre for Scientific Research in Nice, who leads the AIM project.

AIM would have sent a probe, two miniature satellites known as CubeSats and a lander to the asteroid Didymos. There, the scientists planned to watch a NASA craft, known as the Double Asteroid Redirection Test (DART), smash into a 165-metre-wide rock that orbits the asteroid, called Didymos.

ESA had hoped to study how the impact would alter the moon’s orbit. In 2022, Didymos will be close enough to Earth to monitor the deflection from the ground, meaning that DART — which is awaiting its own approval decision in March 2017 — can go ahead without AIM. But the ESA craft would have provided more-detailed knowledge, says Michel.

Wörner left the door open for exploring similar asteroid missions in the future. And ministers supported making a smaller amount of funding available to investigate ways of

taking planetary-protection projects forward, he said.

AIM’s cancellation means that ESA now has no small-body missions planned for at least the next 15 years, a particularly bitter disappointment following the success of ESA’s Rosetta mission, which orbited and landed on a comet. AIM would also have been ESA’s first test of laser communication in deep space, and of its interplanetary CubeSats.

Overall, ESA was able to obtain €10.3 billion of the roughly €11 billion it had requested from its 22 member states, plus Slovenia and Canada, for a variety of programmes. This included funding for the International Space Station up to 2024 and for a future space-weather mission in cooperation with NASA and the Japanese space agency, JAXA.

Ministers also agreed to further finance plans for ESA to work with Roscosmos on the Moon lander Luna 27, a collaboration first pitched in 2014.

ESA will spend around €30 million on preparatory work for the mission, which is scheduled for 2020. European scientists and industry hope to contribute landing, communications and drilling and analysis instruments to the craft, which is designed to study soil and the atmosphere at the Moon’s south pole. ■

ANCESTRY

Ancient genome delivers ‘Spirit Cave Mummy’ to US tribe

Sequencing proves Native American roots of 10,600-year-old skeleton.

BY EWEN CALLAWAY

The sequencing of a 10,600-year-old genome has settled a lengthy legal dispute over who should own the oldest mummy in North America — and given scientists a rare insight into early inhabitants of the Americas.

The controversy centred on the ‘Spirit Cave Mummy’, a human skeleton unearthed in 1940 in northwest Nevada. The Fallon Paiute-Shoshone Tribe has long argued that it should be given the remains for reburial, whereas the US government opposed repatriation. Now, genetic analysis has proved that the skeleton is more closely related to contemporary Native Americans than to other global populations. The mummy was handed over to the tribe on 22 November.

The genome of the Spirit Cave Mummy is significant because it could help to reveal how ancient humans settled the Americas, says

Jennifer Raff, an anthropological geneticist at the University of Kansas in Lawrence. “It’s been a quest for a lot of geneticists to understand what the earliest peoples here looked like,” she says.

The case follows the US government’s decision this year that another controversial skeleton, an 8,500-year-old human known as Kennewick Man, is Native American and qualifies for repatriation on the basis of genome sequencing. Some researchers

lament such decisions because the buried skeletons are then unavailable for scientific study. But others point out that science could benefit if Native American tribes use ancient DNA to secure the return of more remains, because this may deliver long-sought data on the peopling of the region. “At least we get the knowledge

before the remains are put back in the ground,” says Steven Simms, an archaeologist at Utah State University in Logan, who has studied the Spirit Cave Mummy. “We’ve got a lot of material in this country that’s been repatriated and never will be available to science.”

SPIRIT CAVE LAWSUIT

The Spirit Cave Mummy is one of a handful of skeletons from the Americas that are more than 10,000 years old (see ‘Sequencing North American skeletons’). Archaeologists Georgia and Sydney Wheeler discovered it in Nevada’s Spirit Cave in 1940. The skeleton, an adult male aged around 40 at the time of his death, was shrouded in a rabbit-skin blanket and reed mats and was wearing moccasins; he was found with the cremated or partial remains of three other individuals. The Wheelers concluded that the remains were 1,500–2,000 years old. But when radiocarbon dating in the 1990s determined that they were much older, the finds drew

“At least we get the knowledge before the remains are put back in the ground.”

attention from both scientists and the Fallon Paiute-Shoshone Tribe. The tribe considers Spirit Cave to be part of its ancestral homeland and wanted the remains and artefacts.

The US Native American Graves Protection and Repatriation Act (NAGPRA) mandates that remains be returned to affiliated tribes if they are deemed 'Native American' by biological or cultural connections. In 2000, the US government's Bureau of Land Management (BLM), which oversees the land where the mummy was found, decided against repatriation. The tribe sued, and in 2006 a US District Court judge ordered the agency to reconsider the case, calling the BLM's decision "arbitrary and capricious".

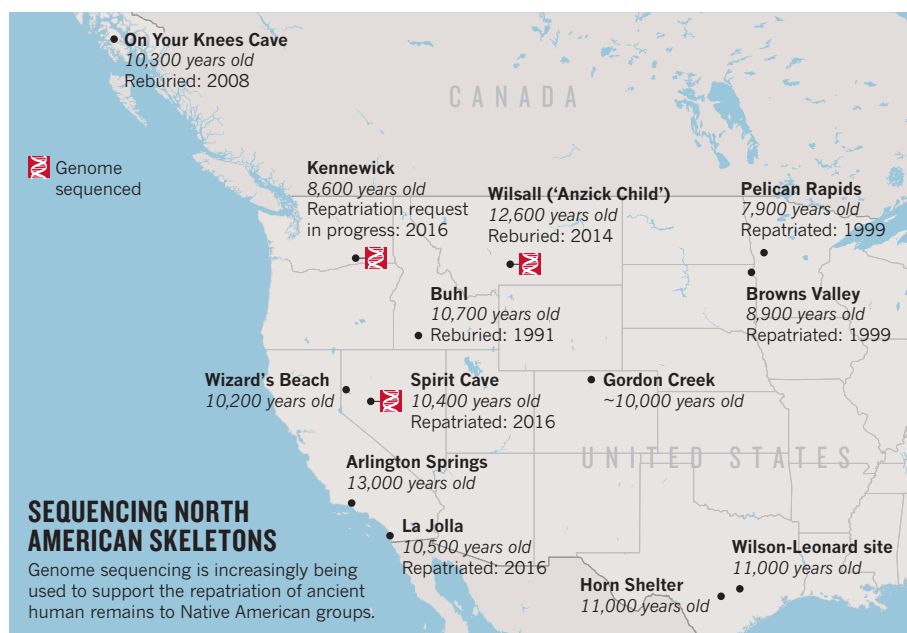
The mummy's remains were stored out of view in a Nevada museum, and placed off-limits to most research, except for efforts to determine its ancestry. In a 2014 monograph based on earlier examination of the remains, US anthropologists Douglas Owsley and Richard Jantz noted that the mummy's skull was shaped differently from those of contemporary Native Americans from the region (*Kennewick Man*, Texas A&M Univ. Press). That contributed to the BLM's decision to seek DNA analysis, says Bryan Hockett, an archaeologist at the bureau's Nevada office in Reno. The tribe was originally opposed to genetic analysis to prove the mummy's ancestry, says Hockett, but eventually agreed.

In October 2015, Eske Willerslev, an evolutionary geneticist who specializes in ancient DNA analysis at the Natural History Museum of Denmark in Copenhagen, travelled to Nevada to collect bone and tooth samples from the mummy and other remains for DNA sequencing, after meeting with tribe members several months earlier.

AMERICA'S EARLIEST PEOPLES

Willerslev's team concluded that the Spirit Cave remains are more closely related to indigenous groups in North and South America than to any other contemporary population. The BLM gave *Nature* a preliminary scientific report from the team, and a 31-page memo outlining its reasoning for repatriating the remains. Willerslev declined to comment because his team's data have not yet been published in a journal.

Hockett says the genome findings offered the only unequivocal evidence that the remains are Native American. No evidence links the remains to any specific group — not even



the ancient DNA — but NAGPRA allows the return of human remains to tribes that have a geographical connection. Len George, chair of the Fallon Paiute-Shoshone Tribe, did not respond to requests for comment.

The genome of a 12,600-year-old skeleton from Montana, called the Anzick Child, is the only other published ancient genome from the Americas that is older than 10,000 years. The Spirit Cave remains and the Anzick Child both seem genetically closer to South American groups than to some North American groups, and the migrations behind this pattern are not yet understood, says Raff. One possibility is that both individuals lived before their local populations began spreading across regions of the Americas, says population geneticist Pontus Skoglund at Harvard Medical School in Boston, Massachusetts.

Sequencing ancient DNA, which has become easier and cheaper in recent years, could help to determine the origins of many other ancient bones. Remains as old as the Spirit Cave Mummy are rare, but there are many younger remains that are not clearly affiliated to any tribe, and which might now be deemed Native American through ancient DNA sequencing and thus repatriated, scientists say.

The BLM announced its intentions to repatriate the Spirit Cave remains in October and

received no formal objections, says Hockett. But Jantz, the anthropologist who co-led the Spirit Cave skull study and is based at the University of Tennessee in Knoxville, laments the decision. "It's just a sad day for science. We will lose a lot of information about the history of human occupation in the Americas as a consequence," he says. Further molecular study of the remains could identify details about the Spirit Cave individuals — from the foods they consumed to the diseases that afflicted them. "I think Willerslev is the last guy who is going to look at these things," Jantz adds.

Dennis O'Rourke, a biological anthropologist at the University of Kansas, says he would like to see more researchers follow Willerslev's example and work with Native American groups to decide whether to sequence ancient human remains, rebury them, or both. And Kimberley TallBear, an anthropologist who studies the views of indigenous groups on genetics at the University of Alberta in Edmonton, Canada, says researchers with O'Rourke's attitude to studying ancient remains are becoming more common. She thinks it is wrong for scientists opposed to repatriation to conclude that tribes are not open to research. "Tribes do not like having a scientific world view politically shoved down their throat," she says, "but there is interest in the science." ■



TOP NEWS



Help wanted:
Canada begins
search for
chief science
adviser
go.nature.com/2gfpkpk

MORE NEWS

- Big biology projects warm up to preprints go.nature.com/2gykczq
- UK moves closer to allowing 'three-parent' babies go.nature.com/2gyloz9
- Science Europe lobby group hit by sudden exodus go.nature.com/2h1aaf0

NATURE PODCAST



The benefits of randomness; fixing brain waves in Alzheimer's; and the DNA of liberated slaves nature.com/nature/podcast

POLICY

Mexico debates ban on human-embryo research

Researchers there have only just started using homegrown human embryonic stem cells.

BY SARA REARDON

Fabián Díaz achieved a milestone last year when he derived the first human embryonic stem-cell line from cells of Mexican origin. Biologists across Mexico now use the stem cells, which Díaz — a researcher at the National Institute of Perinatology in Mexico City — created using embryos discarded by a fertility clinic.

But in recent months, Díaz has put his stem-cell research on hold. He is waiting to see whether Mexico's legislature will approve an amendment to the national health law that would ban experiments with human embryos. The proposal is winding its way through the legislature's lower house, the Chamber of Deputies. To become law, it would have to be approved by the legislature and by Mexico's president, Enrique Peña Nieto.

"They want to eliminate an entire area of research in Mexico," says René Drucker-Colín, a neurobiologist at the National Autonomous University of Mexico (UNAM) in Mexico City who hopes to use embryonic tissue as a treatment for people with Parkinson's disease.

The amendment is intended to regulate assisted reproduction, including the payment of surrogate mothers, donations to egg and sperm banks and the fertilization of more than three eggs at a time. But it would also ban the creation of human embryos for any purpose except reproduction and any research with existing human embryos.

Such restrictions are intended to address Mexico's thriving reproductive tourism industry, which has few protections for surrogate mothers. But the proposed amendment would have prohibited a scientific world first that took place in Mexico: the conception of a baby with DNA from three people. The child was born in April. His parents, who are from Jordan, used the treatment to prevent their baby from inheriting a disease that would otherwise be passed down through his mother's mitochondrial DNA.

The proposed changes to Mexico's health law have the backing of the National Action Party (PAN) and Peña Nieto's Institutional Revolutionary Party (PRI), but researchers worry that they are too broad.

"We're not against the regulation," says Diana Escalante, a neurodevelopmental



MADS NISSEN/BERLINGSKE/PANOS

Reproductive tourism is under threat from proposed changes to Mexico's health law.

biologist at UNAM. "But the way in which they are doing it is just forbidding everything." The amendment would prevent the creation of new embryonic stem-cell lines, she says, as well as a standard way to test whether the cells can develop into any cell type in the body. The penalties for violating the restrictions would include heavy fines and imprisonment.

Human-rights groups have joined scientists in opposing the proposed amendment, which would restrict artificial reproduction to heterosexual couples. Only Mexican-national

couples would be able to use surrogate mothers, who would be limited to their relatives. Opponents of the plan say that it discriminates against same-sex couples and people without family members of reproductive age.

But Rosa Velez, a spokesperson for Sylvana Beltrones, the legislator who authored the amendment, says that the restrictions would protect human dignity and the legal rights of children who are created using fertility techniques and their parents. She adds that scientists would be able to study stem cells

obtained from adults.

Researchers have protested against the plan. On 24 October, more than 60 Mexican scientists sent a letter to the newspaper *El Universal* criticizing the proposed amendment. Drucker-Colín says that he has also asked Mexico's National Academy of Sciences to intercede with the politicians.

EARLY DEVELOPMENTS

The amendment would make it harder for scientists to study the earliest stages of human development, says Iván Velasco, a neurodevelopmental biologist at UNAM and president of the Mexican Society for Stem Cell Research. "It's possible people will train abroad, but if they want to come back they won't be able to do it here," he says. Yet Velasco thinks that his own work, which uses existing human embryonic stem-cell lines, would be permitted.

"It's possible people will train abroad, but if they want to come back they won't be able to do it here."

Others are worried about how a ban on the use of embryonic stem cells would affect clinical research. "We are close to beginning working with [embryonic stem] cells, and these laws are going to

trash everything," says Raymundo Cañales de la Fuente, a research gynaecologist at the Hospital Angeles Pedregal in Mexico City whose group looks for ways to improve the efficacy of assisted reproductive techniques.

The amendment would limit the use of routine techniques used in fertility clinics, including a method used to screen embryos for genetic mutations before they are implanted into the mother. Such screening can prevent the transmission of severe genetic diseases, and help some infertile couples to understand why they are having trouble conceiving.

If the technique is banned, researchers would need to rely on older, less precise methods to determine whether embryos are likely to survive implantation, says Patricia Grether, a geneticist at the National Institute of Perinatology. Clinicians could also send patients to the United States for treatment, but that is too expensive for many Mexicans.

Velez says that the intent of the proposed amendment is to improve assisted reproduction, not to ban it. But Cañales de la Fuente says that the proposal would prevent many reputable clinics from offering such services. Clinicians would be limited to fertilizing three eggs at a time, reducing their success rates. They would also have to verify that a couple is not storing fertilized eggs at another clinic. With more than 100 such clinics in Mexico City alone, there is no practical way to do this.

"We need to make a new law," Cañales de la Fuente says. "Completely different from this one, with a scientific basis and a medical basis to be practical — and from the ministry of health, not from the congressmen." ■

POLITICS

Wage fight leaves US postdocs in limbo

Institutions struggle to respond after court blocks pay law.

BY ANNA NOWOGRODZKI

An ongoing battle over US overtime pay rules has left many postdocs in financial limbo. Labour regulations set to take effect on 1 December would have effectively increased wages for some researchers, but on 22 November a US federal judge in Texas temporarily blocked the rule.

Some universities are proceeding with planned salary increases for postdocs, but others have cancelled — or at least, temporarily halted — changes to researchers' pay. The uncertainty over how the legal fight will play out is already affecting some postdocs' career and family plans.

"The injunction coming down, especially right before the holiday weekend, was really disheartening," says Colm Atkins, a postdoc at the University of Texas Medical Branch at Galveston. His institution had planned increases to comply with the overtime rule, but is now cancelling them. "I know postdocs with spouses and families that were really looking forward to having that safety net."

One couple, both postdocs at the University of Massachusetts Medical School in Worcester, had decided to try for a baby because their combined pay rises would have allowed them to afford childcare, says Sonia Hall, a fellow postdoc at the school. But the institution will not go ahead with the increases, so now they can't, she adds.

NEW THRESHOLD

The US Department of Labor finalized the new wage rule in May. The regulation made overtime pay, set at 1.5 times a worker's hourly wage, mandatory for people making less than US\$47,476 per year, once they work over 40 hours in one week. The average postdoc salary in the United States is about \$45,000 per year, and many researchers surpass the 40-hour cut-off. The previous overtime pay threshold was \$23,660 per year.

Because it would probably have been cheaper and logistically easier to raise salaries than to count the hours a postdoc worked, many universities and government agencies had planned to increase annual wages.

The US National Institutes of Health (NIH) will stick to its plan to raise the minimum salary for postdocs paid through its grants to \$47,484. This is a 9% increase over previous NIH guidelines.

"A lot of places follow the NIH's example," says Kate Sleeth, chair of the board of the National Postdoctoral Association in Washington DC, which has advocated for a minimum postdoc salary of \$50,000 for more than two years. "I'm hoping everyone follows suit."

Even though they are no longer legally compelled to, many institutions — including Duke University in Durham, North Carolina, the University of Minnesota in Minneapolis and Boston University in Massachusetts — will go ahead with plans to raise postdocs' minimum salaries above the \$47,476 cut-off.

For those that aren't raising pay — such as the University of Michigan in Ann Arbor — it is not clear whether they are temporarily pausing their plans or abandoning them entirely. The Future of Research, an advocacy group for junior scientists, is tracking institutional responses to the regulation and its suspension on its website.

Sleeth was not surprised to hear that the regulation had been suspended. "We were kind of waiting to see if someone was going to challenge it," she says. Twenty-one states and a coalition of businesses filed a case against the rule in October, arguing that the Department of Labor had overreached its authority.

"When you're at the bottom of the pecking order, it's hard to ask for a raise."

The judge blocked the rule from taking effect on 22 November, until he could decide on the case. It's unclear when he will do so, or whether the department will challenge the ruling.

"When you're at the bottom of the pecking order, even when you're highly educated and skilled and motivated, it's hard to ask for a raise," says Tess Eidem, a biochemistry postdoc at the University of Colorado Boulder. Her institution will let individual researchers decide whether to raise their postdocs' pay.

The increased wages would have sent a message that postdocs are respected and their work is valued, says Hall. "They're a major driving force of the data collection and discovery in the scientific enterprise, and they want to feel like they're not hiding in the shadows anymore." ■

ENVIRONMENT

Deforestation rates spike in Brazil

Loosening protections damages the Amazon.

BY JEFF TOLLEFSON

Illegal deforestation in the Brazilian Amazon has jumped to its highest level in eight years, raising fears that the country could lose a decade's worth of progress in forest protection.

Satellite data released on 29 November by Brazil's National Institute for Space Research in São José dos Campos show that 7,989 square kilometres of land — nearly the size of Puerto Rico — was cleared between August 2015 and July 2016. That is 29% above the 2015 level, and 75% above the level for 2012 (see 'Going up').

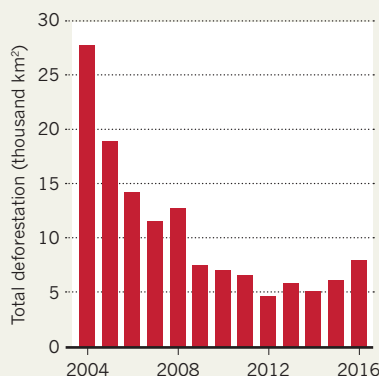
Brazil basked in the international lime-light for nearly a decade after deforestation began to drop in the Brazilian Amazon in 2005, thanks in part to stronger government enforcement.

But Brazil's Congress relaxed the country's deforestation protections in 2012, and there is a push among current lawmakers to further loosen environmental laws. The ongoing economic recession and political-corruption scandals have also diverted money and attention away from environmental enforcement.

This has emboldened ranchers and illegal land traders to resume clearing land, says Paulo Barreto, a senior researcher at the Amazon Institute of People and the Environment, an activist group in Belém. ■

GOING UP

After nearly 10 years of decline, deforestation in Brazil's Amazon is on the rise.



SOURCE: INPE



Marine Le Pen, leader of France's far-right National Front, is challenging the mainstream parties.

SOCIOLOGY

Researchers baffled by nationalist surge

Economic woes wrought by globalization are only part of the cause.

BY JEFF TOLLEFSON

Waves of nationalist sentiment are reshaping the politics of Western democracies in unexpected ways — carrying Donald Trump to a surprise victory last month in the US presidential election, and pushing the United Kingdom to vote in June to exit the European Union. And nationalist parties are rising in popularity across Europe.

Many economists see this political shift as a consequence of globalization and technological innovation over the past quarter of a century, which have eliminated many jobs in the West. And political scientists are tracing the influence of cultural tensions arising from immigration and from ethnic, racial and sexual diversity. But researchers are struggling to understand why these disparate forces have combined to drive an unpredictable brand of populist politics.

"We have to start worrying about the stability of our democracies," says Yascha Mounk, a political scientist at Harvard University in Cambridge, Massachusetts. He notes that the long-running World Values Survey shows that people are increasingly disaffected with their governments — and more willing to support authoritarian leaders.

But this has played out in different ways

across the West. Austria rejected the extreme-right Freedom Party presidential candidate Norbert Hofer in favour of Alexander Van der Bellen, a former Green Party leader, on 4 December. The same day, anti-establishment forces prevailed in Italy, where Prime Minister Matteo Renzi said he would resign after voters rejected his proposed constitutional reforms.

In France, Marine Le Pen has worked to cleanse the far-right National Front of its anti-Semitic roots and build a broader working-class base. Those efforts could make the party, a minor player in French politics since the 1970s, a force during the presidential elections next year.

The bulk of the party's new support is coming from deindustrialized areas in northern France, says Douglas Webber, a political scientist at the global business school INSEAD's French campus in Fontainebleau. "A lot of the industrial workers arguably have made the transition from the extreme left to the extreme right," he says.

In the United States, the regions hardest hit by globalization have become more politically extreme, according to a working paper published in September by David Autor, an economist at the Massachusetts Institute of Technology in Cambridge, and his colleagues. They found that these areas elected more hard-line

ERIC GAILLARD/REUTERS

candidates of both stripes to Congress between 2002 and 2010 — Republicans in majority-white communities and Democrats in ethnically and racially mixed areas. A separate, as-yet unpublished analysis by the team suggests that the trend towards extreme candidates favoured Republicans in presidential elections from 2000 to 2016 — perhaps enough to win Trump the White House this year.

“It was just a matter of time before someone sought to tap into the rich electoral potential inside of a group of people as sizeable as the white working class,” says Justin Gest, a political scientist at George Mason University in Arlington, Virginia. His polling this year suggests that 65% of white US voters would support a hypothetical new protectionist and xenophobic political party. Gest adds that Trump prevailed in part because he gave voters somebody to blame for their economic woes.

Some academics have explored potential parallels between the roots of the current global political shift and the rise of populism during the Great Depression, including in Nazi Germany. But Helmut Anheier, president of the Hertie School of Governance in Berlin, cautions that the economic struggles of middle-class citizens across the West today are very different, particularly in mainland Europe.

The Nazis took advantage of the extreme economic hardship that followed the First

World War and a global depression, but today’s populist movements are growing powerful in wealthy European countries with strong social programmes. “What brings about a right-wing movement when there are no good reasons for it?” Anheier asks.

In the United States, some have suggested that racism motivated a significant number of Trump voters. But that is too simplistic an explanation, says Theda Skocpol, a sociologist at Harvard University. “Trump dominated the news for more than a year, and did so with provocative statements that were meant to exacerbate every tension in the US,” she says.

Trump also prevailed in part because the structure of the US electoral college gives outsized influence to Republican-leaning rural areas over Democratic urban centres. And some of his predominantly white supporters voted for President Barack Obama, a Democrat, in past elections.

OUTLOOK HAZY

Le Pen has called Trump’s win “a sign of hope” for her own campaign, but Webber is sceptical. The latest polls suggest that Le Pen would pick up less than one-third of the vote in a run-off with former prime minister François Fillon, who is running under the banner of the centre-right Republican party, and France’s economy is fairly stable. “I’m not as pessimistic as some,”

Webber says. The rise of nationalism in France has been going on for longer, and seems to have progressed more slowly, than similar surges in many other countries, he says.

Mounk says the broader trend towards nationalism caught political scientists off-guard because they are often too focused on finding new ways to answer mundane questions. Academics must redouble their efforts to understand the nationalist wave and help policymakers to address it, he adds.

“In times of freedom and prosperity, it was nice for us to sit around and pretend to be scientists,” Mounk says. “But right now, if you are twiddling your thumbs with your statistical models instead of thinking about how we can save liberal democracy, you are doing something immoral.” ■

CORRECTION

In the story ‘Speedy drills start hunt for oldest ice’ (*Nature* **540**, 18–19; 2016), the size of the borehole to be drilled by the US RAID was wrong. It will be about 8cm, and so roughly the same size as the other planned holes. The story should also have noted that, as well as Jeffrey Severinghaus, the US project is co-led by John Goodge of the University of Minnesota, Duluth.



FREEDOM IN EXILE

Rupert's Valley is the final resting place for thousands of liberated African slaves. Their DNA could help to unlock the mysteries of one of humanity's darkest chapters.

BY EWEN CALLAWAY

In 1849, Anglican bishop Robert Gray described a slave ship being unloaded on the island of St Helena. "I never beheld a more piteous sight," he observed of the people on board. Some were dead; many more were close to it. "They had a worn look and wasted appearance, and were moved into the boats like bales of goods, apparently without any will of their own."

These men and women were refugees of the British Navy's campaign against the slave trade. The United Kingdom had outlawed the trade in 1807, and anti-slaver patrols were intercepting boats along the Middle Passage — the trade route from Africa to the Americas — even venturing into the harbour of Rio de Janeiro. Situated in the middle of the Atlantic Ocean, St Helena became a favoured drop-off point for people freed from the ships. From 1840 to the late 1860s, scores of ships carrying some 27,000 slaves were captured and brought to the island. The slaves who survived were granted freedom and most were eventually

St Helena, drawn around 1750.

relocated, but for nearly 10,000, many of them children, a rocky valley on St Helena became their final resting place.

These people's lives had been all but forgotten until a decade ago, when a construction project unearthed a couple of skeletons. An archaeological excavation found hundreds more. Now, a team of researchers is studying the remains and sequencing DNA that spent more than a century underground to find out as much as it can about the freed slaves of St Helena: from their birthplaces in Africa to their cultural backgrounds and how they died.

The site's unique place in history could provide important missing information about the experiences and identities of those who were kidnapped but never reached the Americas, says Fatimah Jackson, a biological anthropologist at Howard University in Washington DC, who is sequencing ancient DNA from other slave burial sites. The work has special significance for Jackson, who traces her own heritage to Africa. "We are descendants of people who came through the Middle Passage, American slavery and institutionalized segregation, and in the process lost a lot of details of our identity and our heritage," she says. "Ancient DNA gives us a glimmer into a past we've longed for, but not had direct access to."

VALLEY OF LOST SOULS

Rupert's Valley is a seaside gorge on the remote isle where Napoleon Bonaparte spent his final days in exile. Driving through it, you pass warehouses, a small diesel power station and a fish-processing plant, as well as a few houses and a lone church.

The valley is also dotted with crumbling reminders of its past, such as a metre-wide defensive wall built by the British East India Company and a late-nineteenth-century desalination plant that supported a prison camp during the Second Boer War. But aside from a nondescript stone building, there is little visible reminder of the thousands of former slaves who once resided here.

In November 2006, geologist David Shilston came to St Helena by boat from Cape Town, South Africa — a five-day journey and one of the only ways to reach the island. His employer, the engineering firm Atkins, had been hired to plan an access road for the construction of the island's first airport. Shilston's job was to explore the geology beneath it. The digging had just begun when a worker called him over, pointing out a broken human leg bone sticking out of the ground. "I thought it might have been a child. It looked pretty small," Shilston says.

Historical records illustrate the dire condition in which many newly liberated slaves arrived. "Their arms and legs were worn down to about the size of a walking-stick," wrote one witness of an 1861 slave-ship landing. "Many died as they passed from the ship to the boat ... there was no time to separate the dead from the living." Diseases such as smallpox and dysentery killed scores more soon after they landed, and there are oblique references to suicides. Between 1840 and 1849, one-third of the freed slaves, nearly 5,000 people, died. Many of those who survived the camp's miserable conditions were sent to Jamaica, Trinidad and other British colonies, where they served as indentured workers on sugar plantations.

After Shilston found the first remains of St Helena's Liberated African Graveyard, as the site is now known, he learned that the burials hadn't been so much forgotten as ignored by many of St Helena's residents. His discovery made the front page of a local newspaper. "There is no doubt that numerous remains of human bodies will be found," the story read. "Many believe that the souls of the dead slaves still are haunting Rupert's." The remains were stored in small caskets at a local church, and later reburied in a cemetery in the island's capital, Jamestown.

"ANCIENT DNA GIVES US A GLIMMER INTO A PAST WE'VE LONGED FOR, BUT NOT HAD DIRECT ACCESS TO."

The UK government's Department for International Development was footing the £250-million (US\$310-million) bill for the airport (St Helena is a British Overseas Territory). The department wanted the site professionally examined, so in 2007 it sent over Andrew Pearson, a commercial archaeologist based in Cardiff, UK. He identified two distinct burial locations, smack in the middle of the path for the planned road. "I was the bearer of the bad news that they had a large and internationally unique burial right where they wanted to build," he says.

Given instructions to excavate only the skeletons beneath the planned road, Pearson and his colleague Ben Jeffs returned to St Helena the following year with a team of 5 people; it quickly swelled to 15 when the size of the task became clear. The first graves that the group excavated were densely packed together, and some contained as many as six skeletons. Pearson's team spent 4 months exhuming 325 individuals from a 1,800-square-metre plot and boxing them up for later examination in Jamestown. More than half looked as if they were under 18, and the largest demographic group was children aged 12 and younger. Pearson estimates that a total of 8,000 people were buried across the entire valley.

North and South America and the Caribbean islands are full of burial sites related to the transatlantic slave trade, but these tend to hold multiple generations of slaves: some born in Africa and others in the Americas. Rupert's Valley, by contrast, contains physical evidence of the individuals who were sold into slavery and then freed. It therefore has the potential to indicate not only where slaves came from and their condition aboard the ships, but also how the slave trade might have reshaped their identities. "It's literally people who are kidnapped in Africa, weeks before," says Pearson. "So you actually have a snapshot of the Middle Passage, which in other ways is completely intangible."

HIDDEN IDENTITY

Hannes Schroeder, an ancient-DNA researcher at the University of Copenhagen, first encountered archaeological remnants of the slave trade during an excavation in the Virgin Islands as an undergraduate. The site he was working on dated to before the arrival of Europeans in the Americas, but it happened to abut an overgrown eighteenth-century plantation. He and a friend spent their weekends investigating the plantation, and identified the remnants of the great house, slave quarters and other landmarks. "There was this site, forgotten, but it was important in terms of the history of the island," he says. "That had a very profound effect on me."

In 2009, as he was starting a postdoc in Copenhagen, Schroeder started to think about determining where slaves had actually come from by looking at their genomes. The history of the roughly 12 million Africans kidnapped as part of the slave trade is based largely on maritime records — including a database of nearly 36,000 voyages. These are largely transactional and generally mention only a slave's port of departure from Africa, not their ethnicity or geographic origins, which could lie hundreds of kilometres from where they were sold. But DNA collected from remains and matched to a genome database of modern Africans could, Schroeder realized, link victims of the slave trade to their homelands.

It was a challenging proposition for two reasons. First, although scientists have recovered readable bits of DNA dating back hundreds of thousands of years, most of the success has come from sites in northern Europe, Siberia or North America. No one has had much luck sequencing ancient DNA from tropical locations, because heat and humidity hasten the breakdown of biomolecules. Second, the genetics community has historically failed to fully represent the genomic diversity of modern

Africa in databases. Even if high-quality sequences could be obtained from slave burial sites, it would be difficult to link them to contemporary populations with much precision.

But Schroeder persevered. In 2011, he and his colleagues won a €4-million (US\$5.8-million) grant from the European Commission to apply scientific approaches to the study of the transatlantic slave trade, and he and others are now sequencing remains from half a dozen slave burial sites, including those on St Helena.

The first bones that the team analysed were from three individuals discovered by construction workers in the Dutch territory of Saint Martin in the Caribbean. Carbon dating suggested that the two men and one woman died between 1660 and 1688, when records showed at least one slave ship arriving at the island. All three had their incisors filed or chipped to points, hinting that they were born in Africa, where tooth modification was a common rite of passage among many ethnic groups. Using molecular techniques that 'fish out' bona fide ancient human DNA from the microbial genetic material that invades bones after death, Schroeder's team managed to recover partial genome sequences for each person.

Comparisons with DNA from 11 contemporary African groups suggested that one of the St Martin individuals was most closely related to members of a Bantu-speaking ethnic group in northern Cameroon called the Bamoun, whereas the other two had DNA in common with ethnic groups that now live in Nigeria and Ghana, including the Igbo and Brong. Schroeder is hesitant to conclude that they belonged to these groups, but the findings, published last year (H. Schroeder *et al. Proc. Natl Acad. Sci. USA* **112**, 3669–3673; 2015), gave him hope that ancient DNA could offer genuine insights into the genetic ancestry of slaves, as well as how the Middle Passage reshaped their identities. "These three individuals — despite the fact that they were found buried together and may have arrived on the same vessel — had different ethnic backgrounds" and probably spoke distinct languages, he says. "That makes you think: how did they communicate with each other and what does this mean for the formation of new identities in the Americas?"

Rupert's Valley was probably even more polyglot. Shipping records suggest that slave ships that landed in St Helena had embarked from ports across Central and West Africa, including present-day Angola and the Congo region (see 'The route to Rupert's Valley'). There are indications that some individuals came from much further away, including

Mozambique and even Madagascar. One 1840s observer noted a group of 40 "natives of the interior", who had travelled for several months to reach the Atlantic coast. "They are of so many different tribes and districts that it would be curious, if one knew the languages, to trace them out," said another visitor to Rupert's Valley.

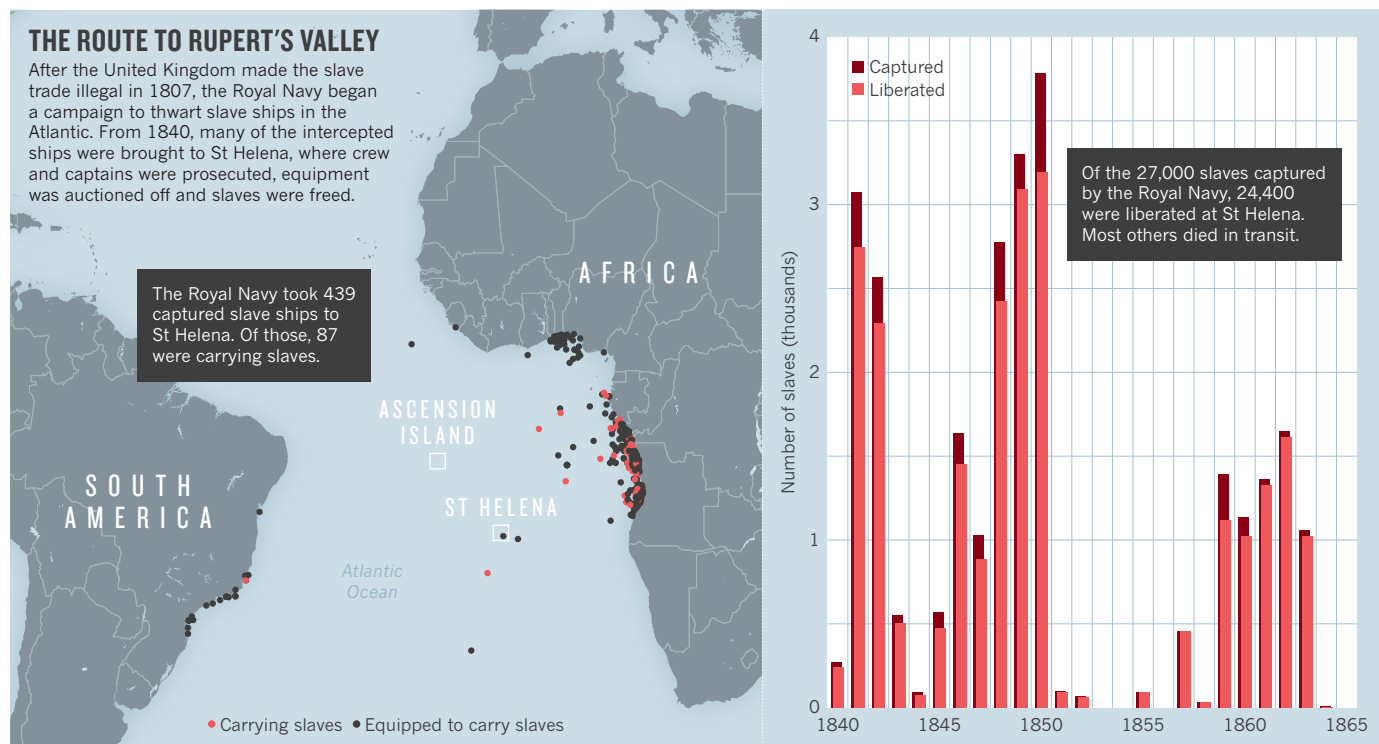
The former slaves' DNA backs up the observation. Schroeder and Marcela Sandoval Velasco, a palaeogeneticist at the University of Copenhagen, collected DNA from the teeth of 63 individuals and sequenced partial genomes from 20 of the best-preserved samples. Comparisons with contemporary African populations suggested that the liberated slaves came from diverse African backgrounds. A few individuals shared ancestry with contemporary West and Central African ethnic groups such as the Bamoun and Kongo, but for most, none of the African groups that the team compared them with was an especially close match. Schroeder attributes that to a lack of genomic data from places such as Angola and Mozambique.

ALTERNATIVE EVIDENCE

Schroeder is frustrated that his team cannot yet pinpoint where the liberated slaves of St Helena were from, nor link them closely to any contemporary population. But he is confident that, eventually, their genomes will help to narrow down that search. Several biomedical projects are sequencing the genomes of lots of people from across sub-Saharan Africa. "It won't take long," he says.

Until then, he and his colleagues are following other leads to trace the origins of the 325 liberated slaves and — equally important, they say — to fill in details about their lives. Levels of certain chemical isotopes, including strontium, in teeth vary depending on local geochemistry, offering clues to where individuals lived as children, when their adult teeth were formed.

Judy Watson, a PhD student at the University of Bristol, UK, who is collaborating with Schroeder's group, has found that groups of individuals in Rupert's Valley had similar isotope levels, as if they had grown up in the same area. "That could be evidence for people being picked up, not just as individuals, but as groups of people from the same region," says Kate Robson-Brown, a biological anthropologist who leads the Bristol group. And if genome analysis can link the liberated slaves to broad modern populations, isotope levels — which can vary enormously between nearby



SOURCE: A. PEARSON ET AL. INFERNAL TRAFFIC (COUNCIL FOR BRITISH ARCHAEOLOGY, 2011)



One of the liberated slaves buried in Rupert's Valley. Few of the bodies were found in coffins; most were buried directly in the earth.

areas — could narrow down their geographical origins. With Pearson's help, Watson has even drawn up a list of potential slave ships and ports of departure, based on coins and other archaeological evidence that suggest the individuals recovered from Rupert's Valley died in the 1840s.

In the quest to learn as much as possible about the lives of the former slaves, the team is also looking at other lines of evidence. "Isotopes and genetics are not going to tell us about the social identity of that person," says Robson-Brown.

For instance, a graduate student in her lab, Erna Johannesdottir, is examining the variety of cultural tooth modifications seen in the remains. Specific modifications could hint at the bearers' ethnic affiliations, especially when compared with the genome data. Some of the modifications seem to have been performed not long before the individuals died, Robson-Brown notes, possibly even aboard the slave ship. "It's not impossible to imagine that these young people are developing kinship groupings among themselves during this process of capture, enslavement and shipping and all these awful things. That, to me, is such a strong message of the human spirit and resilience."

Scholars of the transatlantic slave trade are eager to weave these insights into their research, says David Richardson, an economic historian at the University of Hull, UK, who is working with the team. "Historians are fascinated by issues of identity and also by issues of cultural adaptation and cultural legacies. This is all part of the transatlantic slave trade."

ANCESTRAL LINKS

DNA research into the slave trade may also reshape how people think about the trade's legacy. "It's a very personal thing for us," says Jackson. "It's not just ancient DNA. It's ancient DNA of potential ancestors." Even if it never becomes possible to link former slaves to their living descendants through DNA, the mere presence of genetic material holds power, Jackson says. "Ancient DNA research allows us to connect with history, both the good and bad, and reconcile it with our own identity."

Sociologist Alondra Nelson at Columbia University in New York City says that this kind of genetic root-seeking started around 1991, with the discovery of a slave burial site in lower Manhattan thought to hold between 10,000 and 20,000 individuals. Mitochondrial DNA analysis (some of it by Jackson) could not say much about the individuals recovered from the site, except that they carried maternal markers common in Africa. But even years before this tentative genetic link was made, people claiming that they had connections to the buried individuals had come forward. A group called Descendants of the African Burial Ground criticized politicians and researchers for mishandling the remains.

Pearson worried that the excavations in Rupert's Valley would draw a similar response from the island's residents (known locally as Saints),

many of whom are descended from earlier slave populations, the Chinese indentured workers who replaced them, and British settlers.

"I thought we'd have terrible trouble — descendants coming up asking, 'How are you doing this to my people?'" he says. But no one ever came forward. "I think people probably chose not to emphasize their black heritage, and it's just been lost."

Although St Helena's airport has been completed, its wind-blasted runway has yet to receive regular commercial flights — and it may never do so. The fate of the remains of the liberated slaves is similarly in question. In 2015, St Helena's government conducted a survey of residents, asking what to do with the skeletons. Options ranged from maintaining them as a research resource to repatriating them to a yet-to-be-determined location in Africa (a proposition that, even if their origins could be pinpointed, would be filled with political and logistical hurdles, Pearson notes). The vast majority wanted them put back to rest in Rupert's Valley — and many indicated that it should be done as soon as possible.

As much as researchers could still learn from the bones, Pearson agrees with the will of the Saints, and hopes to return to St Helena to help with reburial. "These people have had such an awful end-of-life experience. What final indignity at the hands of white people to be stuck in a lab for perpetuity."

In 1864, Rupert's Valley received its last six liberated slaves from a ship so damaged that the Royal Navy left it to sink. The refugee camp was disbanded in 1867.

Historians are ambivalent about the campaign to intercept and prosecute slave traders. It probably hastened the end of the transatlantic slave trade in Brazil and Cuba, the last holdouts of a dark chapter in human history. But the people liberated by the British Navy who eventually left St Helena may have considered their fates — working in foreign lands or fighting other people's wars — merely an extension of their enslavement.

Then there were those who remained. Several hundred liberated Africans eventually integrated into St Helena's population. An 1881 census recorded 77 people whose birthplace was listed as "West Coast of Africa."

Their lives are not easy to imagine — ripped from their homelands, saved from enslavement but left on a remote island. Free but exiled. One turn-of-the-century photograph shows five of them: three seated women and two men, all residents of the poorhouse. "The men, although over seventy, are still able to earn a little, but the women are helpless, and almost blind," reads a caption from the time. They had arrived on the *Cyclops*, a naval vessel that last visited St Helena half a century earlier. It is the final picture ever taken of St Helena's liberated slaves, but DNA may soon offer one more. ■

Ewen Callaway writes for Nature from London.

COMMENT



SOCIAL SCIENCE Take a broader, longer view of the Anthropocene **p.192**

CHILDREN The bedtime stories that set *Nature's* editors on the road to science **p.194**

CONSERVATION Halt catastrophic extinctions in Brazil's Cerrado **p.199**

COMMUNITY Technicians need career paths, metrics and rewards **p.199**

ADRIANE OHANESIAN/REUTERS



A mother and her child at a hospital in South Sudan, one country where precise public-health data are lacking.

Four steps to precision public health

Better surveillance data and analyses are urgently needed to control disease in the developing world, argue **Scott F. Dowell, David Blazes and Susan Desmond-Hellmann**.

When domestic transmission of Zika virus was confirmed in the United States in July 2016, the entire country was not declared at risk — nor even the entire state of Florida. Instead, precise surveillance defined two at-risk areas of Miami-Dade County, neighbourhoods measuring just 2.6 and 3.9 square kilometres. Travel advisories and mosquito control focused on those regions. Six weeks later, ongoing surveillance convinced

officials to lift restrictions in one area and expand the other.

By contrast, a campaign against yellow fever launched this year in sub-Saharan Africa defines risk at the level of entire nations, often hundreds of thousands of square kilometres. More granular assessments have been deemed too complex.

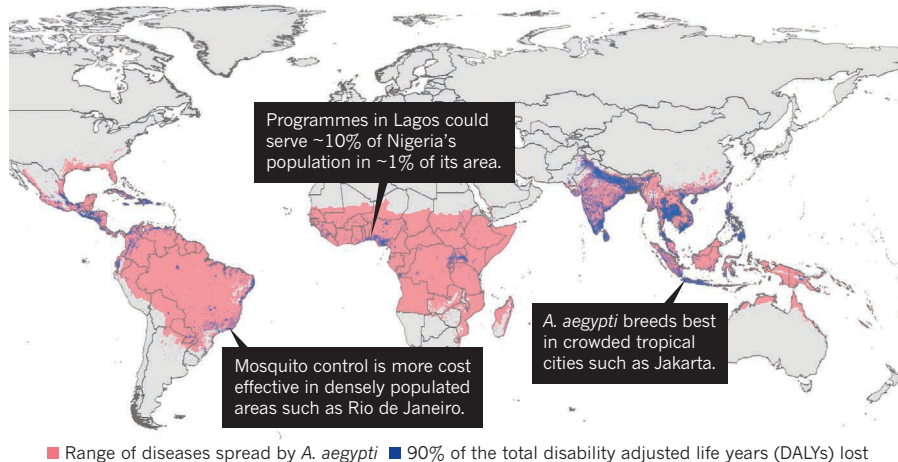
The use of data to guide interventions that benefit populations more efficiently is a strategy we call precision public health. It requires

robust primary surveillance data, rapid application of sophisticated analytics to track the geographical distribution of disease, and the capacity to act on such information¹.

The availability and use of precise data is becoming the norm in wealthy countries. But large swathes of the developing world are not reaping its advantages. In Guinea, it took months to assemble enough data to clearly identify the start of the largest Ebola outbreak in history. This should take ►

STRATEGIC DEFENCE

Ninety per cent of the disease burden can be addressed by focusing on just 14% of the total area in which the mosquito *Aedes aegypti* transmits chikungunya, dengue, yellow fever and Zika.



► days. Sub-Saharan Africa has the highest rates of childhood mortality in the world; it is also where we know the least about causes of death.

We at the Bill & Melinda Gates Foundation in Seattle, Washington, are committed to seeing data used equally in the developing world. The United Nations' Sustainable Development Goals, launched in September 2015, will require ever more accurate and timely data to track and achieve progress. We are calling on government leaders and global health workers to join us in revamping surveillance to make precision public health a reality.

DATA TRENDING

The value of precise disease tracking was baked into epidemiology from the start. In 1854, John Snow famously located cholera cases in London. His mapping of the spread of infection through contaminated water dealt a blow to the idea that the disease was caused by bad air. These days, people and pathogens move across the globe swiftly and in great numbers. In 2009, the H1N1 'swine flu' influenza virus took just 35 days to spread from Mexico and the United States to China, South Korea and 12 other countries.

Fortunately, data are also more available than ever before. The Global Polio Eradication Initiative routinely sequences all poliovirus isolates collected from people diagnosed with the disease. The presence of viruses with relatively large (>1.5%) sequence differences from other isolates can indicate where surveillance and vaccination efforts have failed, and suggest how to redirect efforts to reduce undetected transmission².

Large-scale collections of the world's health data are catalogued (such as by the Global Health Data Exchange: <http://ghdx.healthdata.org>) and have been used to compile an inventory of the global burden of disease. The 2015 report from the Institute for Health Metrics and Evaluation in Seattle

enlisted more than 1,800 collaborators who analysed more than 14,000 data sets to estimate mortality for 249 causes in 195 countries across 35 years³.

The public-health community is sharing more data faster; expectations are higher than ever that data will be available from clinical trials and from disease surveillance. In the past two years, the US National Institutes of Health, the Wellcome Trust in London and the Gates Foundation have all instituted open data policies for their grant recipients, and leading journals have declared that sharing data during disease emergencies will not impede later publication (see go.nature.com/2gcwh4a).

Meanwhile, improved analysis, data visualization and machine learning have expanded our ability to use disparate data sources to decide what to do. A study published last year⁴ used precise geospatial modelling to infer that insecticide-treated bed nets were the single most influential intervention in the rapid decline of malaria.

However, in many parts of the developing world, there are still hurdles to the collection, analysis and use of more precise public-health data. Work towards malaria elimination in South Africa, for example, has depended largely on paper reporting forms, which are collected and entered manually each week by dozens of subdistricts, and eventually analysed at the province level. This process would be much faster if field workers filed reports from mobile phones⁵.

SPEED, ACCURACY, EQUITY

Precision public health can save lives. Meningococcal disease in West Africa can go from scattered cases to devastating outbreaks in weeks. Since 1997, regular surveillance has triggered mass vaccination

campaigns when clusters of disease surpass set thresholds. A 2015 analysis⁶ of surveillance data in Niger found that monitoring at the level of 'health areas' (median population of 14,440), rather than at the district level (median population of 295,200), would increase the signals detected and reduce the delay in response. Accelerating effective vaccination by just two weeks would interrupt any outbreak more quickly, halving the number of cases. However, the analysis also showed that such outcomes would require more-frequent data analysis and decentralized vaccine storage.

More-precise disease surveillance can also illuminate causes of disease and spark opportunities for prevention. In the pilot phase of a new Gates Foundation surveillance programme, investigators in Soweto in South Africa found evidence of group B streptococci in the tissues of stillbirths (see 'Disease surveillance'). Women who have experienced such losses also have a seven-fold higher risk of recurrence in their next pregnancy. Penicillin can prevent group B streptococcal infections in newborns, and prescribing the antibiotic to women with risky pregnancies is feasible and affordable for poor communities such as Soweto — but only if health workers know how to target treatment to the fraction of patients who need it. Thus, improved surveillance and analysis allows an impoverished population to receive an intervention that is standard practice in wealthy countries, but which is not used in much of the developing world.

Geographic precision also means that public-health resources are used more efficiently, so more people receive interventions. For example, the *Aedes aegypti* mosquitoes that transmit dengue, Zika and chikungunya viruses do not spread these infections if they carry *Wolbachia*, a benign symbiotic bacterium⁷. *Wolbachia*-carrying mosquitoes can be distributed to displace wild-type *A. aegypti* and reduce transmission of these viral diseases, but the wide global prevalence of these arboviruses makes this approach daunting and potentially costly⁸.

Fortunately, disease-distribution information and geospatial modelling has pinpointed the areas that have 90% of the predicted disease burden (see 'Strategic defence'). This reduces the area for intervention from some 50 million square kilometres to 7 million square kilometres, and promises to make *Wolbachia*-based mosquito control feasible and affordable for many more countries. Demonstration projects in Medellín in Colombia and Rio de Janeiro in Brazil began in late 2016.

Still, we must try harder. Despite progress, data quality is often too poor to drive decisions. Data sets are limited, testing is unreliable and outcomes from vast regions

are unavailable. In 2013, errors in Google Flu Trends, an open flu-surveillance tool, showed that big-data analytics alone can supplement but not replace conventional surveillance⁹. Furthermore, the cause of most deaths is never adequately identified — in 2008, only 4% of childhood deaths worldwide occurred in countries that have an adequate medical certification procedure.

ACHIEVING PRECISION

Four concrete steps are necessary for precision public health to become regularly available in the developing world.

Register births and deaths. It is hard to know whether a national deworming programme for children in Sierra Leone or a vaccination programme for whooping cough (pertussis) in Nigeria is reducing mortality when less than 4% of deaths are registered.

These basic demographic data are necessary for rational public-health decisions. To achieve the Sustainable Development Goals, the World Bank Group and its partners should expand support for systems to sample representative areas for births and deaths in all high-mortality countries by 2020, as a step towards full civil registration and vital statistics in all countries by 2030. Our estimates put this cost at less than US\$1 million per year for countries with 30 million people or fewer.

Track disease. Careful surveillance can guide public health in a country and track disease outbreaks that could spread beyond borders. This requires infrastructure and systems to collect and analyse data,

laboratories to confirm diagnoses, and adequately trained personnel. All 194 member states of the World Health Organization (WHO) should fulfil their obligations under the International Health Regulations to achieve a minimum core capacity for surveillance. When countries are too poor for this, G20 countries should help. The estimated cost of pandemic preparedness is about \$4.5 billion — less than 0.005% of the world's gross domestic product.

At the 2017 World Health Assembly in May, these rich countries should commit to implementing concrete support for at least 76 of the neediest ones. Countries should submit to objective, published, external evaluations of their surveillance capacity and pandemic preparedness; self-reports of progress are not acceptable. The Joint External Evaluations, led by the WHO and building on targets developed under the Global Health Security Agenda, should guide this process.

“The cause of most deaths is never adequately identified.”

Incorporate laboratory analyses. The cause of a child's death cannot be adequately determined by interviews with relatives, especially months after death. Where feasible, tissue sampling and laboratory diagnosis should supplement or replace these ‘verbal autopsies’. Our initial experience is that most relatives consent readily, because families want to know why their child died. Autopsy-based mortality surveillance is expensive — more than \$5 million per country per year

— but can strengthen broader surveillance efforts. By pairing just a small number of physical autopsies with verbal autopsies on the same deaths, the much larger number of verbal autopsies can be calibrated to more accurately reflect the true distribution of mortality causes.

Train more people. Public-health personnel at national and local levels must be educated in the fundamentals of epidemiology, empowered to use local information to set strategies, and equipped to translate those decisions into action.

Relevant resources exist: field epidemiology and laboratory training programmes such as FETP, supported by the US Centers for Disease Control and Prevention, are highly effective. Software such as Epi Info 7 provides basic analysis for outbreak management, and other open-source software products assist with disease surveillance, such as the Suite for Automated Global Electronic bioSurveillance (SAGES) from Johns Hopkins University. However, these are not deployed in many places where they are desperately needed.

Training costs are not high — we estimate that salaries would be needed for perhaps a dozen doctoral-level personnel per country, but even wealthy countries struggle to maintain budgets for public-health workers when there is no crisis to put pressure on politicians. The WHO and other players must keep the pressure on.

Frontline workers should not find themselves frustrated by global programmes that fail to take into account data on local circumstances. Wherever they live — in a village, city or country, in the global south or north — people have the right to public-health decisions that are based on the best data and science possible, that minimize risk and cost, and maximize health in their communities. ■

DISEASE SURVEILLANCE

Why investment is key

The Bill & Melinda Gates Foundation did not initially focus on disease surveillance; this is conventionally a government function. Yet it became obvious that the lack of data prevented us from modelling disease or designing effective vaccine trials.

In 2014, we established a team focused on surveillance and epidemiology. Its first task was tracking childhood mortality using a technology called minimally invasive tissue sampling. This provides more precise information than interviews with relatives about deceased children's symptoms.

These efforts have grown into a network of sites across Africa and South Asia called Child Health and Mortality Prevention Surveillance (CHAMPS). The sites are geographically limited, with well-defined populations of 150,000 to 300,000 people. Active outreach by community health workers helps to identify pregnancies and

deaths of children under five, and to obtain tissue samples for definitive identification of the causes. CHAMPS will now be integrated with efforts to track mortality countrywide in Mozambique and elsewhere.

Pathology findings from CHAMPS sites will be combined with ongoing programmes from the World Bank, UNICEF, the World Health Organization, Canada, the US Agency for International Development and the US Centers for Disease Control and Prevention to collect vital national statistics. The Gates foundation plans additional investments in these programmes. This information will be fed into the data sets accessed by the Institute for Health Metrics and Evaluation in Seattle, Washington, and others.

This will expand mapping capacity and improve estimates of the global distribution of disease and the impacts of public-health interventions.

Scott F. Dowell is deputy director for surveillance and epidemiology; **David Blazes** is senior programme officer for surveillance and epidemiology; and **Susan Desmond-Hellmann** is chief executive at the Bill & Melinda Gates Foundation, Seattle, Washington, USA.
e-mail: scott.dowell@gatesfoundation.org

1. Desmond-Hellmann, S. *Science* **353**, 731 (2016).
2. Wassilak, S. G. F. *et al. J. Infect. Dis.* **210** (Suppl. 1), S5–S15 (2014).
3. GBD 2015 Mortality and Causes of Death Collaborators. *Lancet* **388**, 1459–1544 (2016).
4. Bhatt, S. *et al. Nature* **526**, 207–211 (2015).
5. Quan, V., Hulth, A., Kok, G. & Blumberg, L. *Malaria J.* **13**, 151 (2014).
6. Mainassara, H. B. *et al. Emerg. Infect. Dis.* **21**, 1322–1329 (2015).
7. Ferguson, N. M. *et al. Sci. Transl. Med.* **7**, 279ra37 (2015).
8. Kraemer, M. U. G. *et al. eLife* **4**, e08347 (2015).
9. Olson, D. R., Konty, K. J., Paladini, M., Viboud, C. & Simonsen, L. *PLoS Comput. Biol.* **9**, e1003256 (2013).



ALEXGCS/GETTY

Terraced rice fields in China: agriculture is one of the many human influences that has shaped Earth's history.

Involve social scientists in defining the Anthropocene

The causes of Earth's transition are human and social, write **Erle Ellis** and colleagues, so scholars from those disciplines must be included in its formalization.

Three dozen academics are planning to rewrite Earth's history. The Anthropocene Working Group of the International Commission on Stratigraphy (of which one of us, E.E., is a member) announced in August that over the next three years it will divide Earth's story into two parts: one in which humans are a geological superpower — an epoch called the Anthropocene — and the other encompassing all that came before our species had a major influence on Earth's functioning¹.

Where to put the transition is being debated. Discussions have narrowed to defining one or more 'golden spikes': sharp global signatures in the rock record derived from the introduction of mid-twentieth century technologies, from radionuclides to plastics. Such markers will be put forward as the basis for ratifying the epoch by the International Geological Congress.

We agree that human influences on the planet should be recognized — but the formalization of the Anthropocene should not be rushed. And we question the privileging of 1950s-era markers. This ignores millennia of previous human influences, from our

use of fire to the emergence of agriculture^{2–6}. Moreover, these markers misrepresent the continuous nature of human changes to our planet. They instil a Eurocentric, elite and technocratic narrative of human engagement with our environment that is out of sync with contemporary thought in the social sciences and the humanities^{3,7–9}.

Decades of rigorous scientific research into the history, causes and consequences of the long-term reshaping of Earth systems by humans is being ignored in the group's discussions. How can a human-centred geological period be defined without characterizing the development of societies, urbanization, colonization, trading networks, ecosystem engineering and energy transitions from biomass to fossil fuels?

We call for the Anthropocene formalization process to be rebuilt on a rigorous, transparent, open and sustainable foundation in which the human sciences have a major role.

DEEPER AND THICKER

The Anthropocene was not made in a day, nor was it created uniformly: the material records of human alterations of Earth

are thick, deep and heterogeneous. They highlight huge social, cultural and technological differences across time and space^{7,8}.

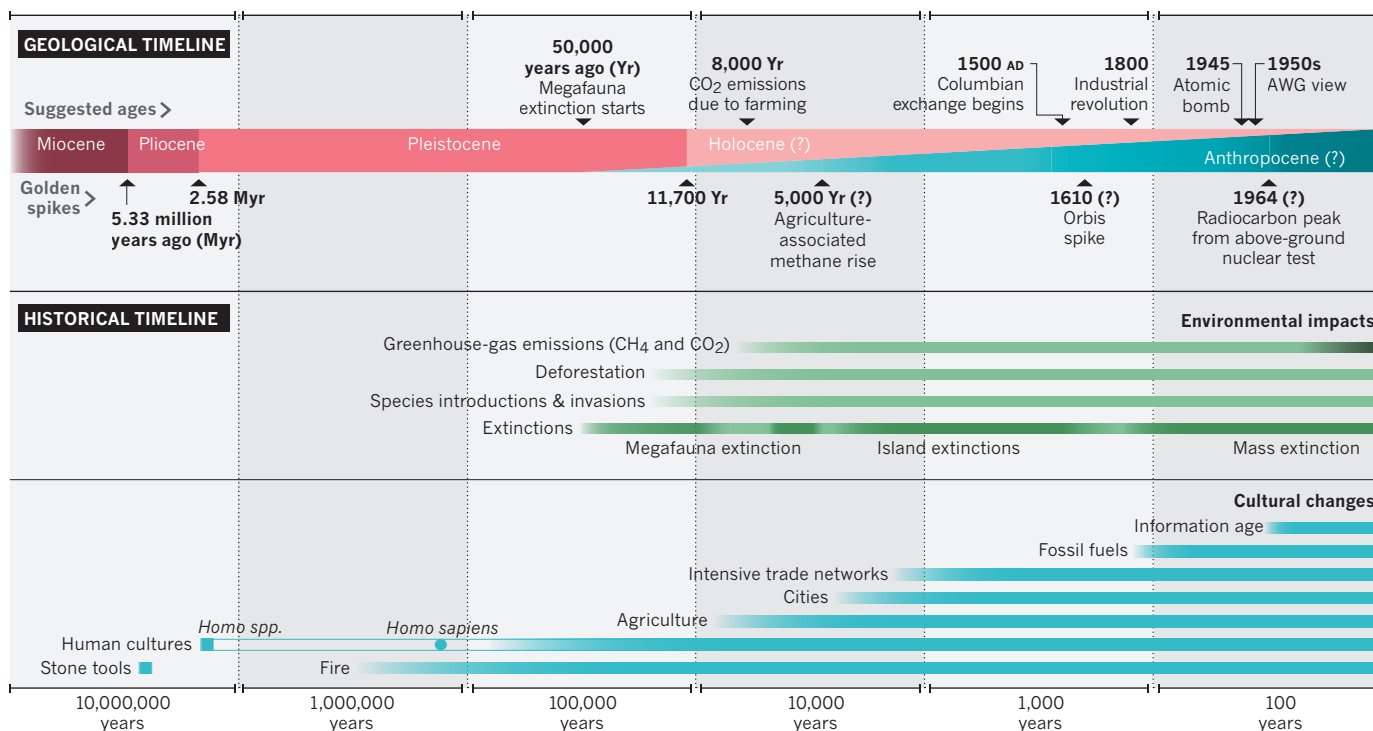
Human activities over the past 10,000 years have caused extinctions and global changes in the distribution of wild and domesticated plants, animals and microflora. Land clearance has altered patterns of erosion and released greenhouse gases into the atmosphere. Humans have created materials such as ceramics, brick and concrete as well as pollutants. Vast networks of canals, reservoirs and irrigation — such as those associated with the Angkor Wat temple complex in Cambodia — have shaped lands and ecologies^{2–4,6,10}.

Agriculture, which emerged in more than a dozen places at different times starting more than 10,000 years ago, has left a vast and indelible record across most of Earth's continents. Although no one yet knows the fate of plastics, the fossil record of agriculture is well documented in ancient pollen, seeds, parasites, bones, deposits of charcoal and soils. Giant irrigation networks can be traced from the air or space.

Earth sciences long ago moved away from

THE DEEP ROOTS OF THE ANTHROPOCENE

Human societies began altering Earth long ago. Human social and cultural capacities to alter its environmental processes have accumulated, scaled up and reinforced each other in complex and historically contingent ways. Defining an Anthropocene epoch should involve examining these transformative social-environmental changes, rather than solely focusing on globally instantaneous environmental transitions. 'Golden spikes' mark stratigraphic boundaries of geological time periods; '?' highlight recent boundary proposals.



AWG, Anthropocene Working Group

defining precise stratigraphic boundaries to developing records of continuous change¹. Isotope 'proxies' trace the rise and fall of global temperatures, ice volumes and atmospheric gases. Earth-systems models link together slow shifts in atmospheric carbon, sea levels and isotopes in seawater and marine deposits. Likewise, agriculture, trade and industrialization are gradual processes that emerged at different times across Earth (see 'The deep roots of the Anthropocene').

Understanding 'human systems' requires engaging a vast body of scholarship based on a diverse array of records (including archaeological, historical and palaeoecological) and perspectives (from political ecology, political economy, historical ecology, cultural evolution and environmental ethics, for instance). Understanding changes in global climate, for example, requires knowing how social and cultural processes drive the clearance of agricultural land and exchanges of atmospheric greenhouse gases, moisture and energy. These processes range from the practices of agricultural land management to demographic shifts, land grabbing and societal conflict.

The Anthropocene Working Group has thrown in a few deeper anthropogenic signals, such as pollution caused by the first production of metals. But these have hardly been considered because the records vary in extent, timing and geographical availability. Instead, the group has focused

almost exclusively on geological deposits that pinpoint one event simultaneously around the world. The reason seems clear to us. Although the group does include members outside the natural sciences (such as a journalist, a lawyer and historians of science) only 3 of the 37 members are social scientists who study long-term social change (two archaeologists and one historian).

MORE INCLUSIVE

The formalization of the Anthropocene must be more transparent and have wider input and assessment. The criteria for assessing the sciences of the new epoch need to be published and peer reviewed, rather than agreed in private meetings. An open online platform could host the full range of proposals and research papers as well as feedback and discussion. The Intergovernmental Panel on Climate Change and the UK Royal Society and US National Science Foundation Assessment Reports serve as models.

A dedicated scientific institution, perhaps called the International Anthropocene Commission, should coordinate this. It could be set up and funded under the auspices of the International Geological Congress, Future Earth (a ten-year international research initiative on global change) and the United Nations. Half of its members should be drawn from anthropology, archaeology, history, sociology, geography, palaeoecology,

economics and philosophy. It should have a formal procedure for inclusion.

Defining a human-centred epoch will take time. It should be treated by scholars from all disciplines with the seriousness it deserves. ■

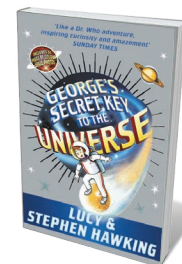
Erle Ellis is in the Department of Geography & Environmental Systems, University of Maryland, Baltimore County, Baltimore, Maryland, USA. **Mark Maslin** is in the Department of Geography, University College London, UK. **Nicole Boivin** is at the Max Planck Institute for the Science of Human History, Jena, Germany. **Andrew Bauer** is in the Department of Anthropology, Stanford University, Stanford, California, USA. e-mail: ece@umbc.edu

1. Waters, C. N. *et al. Science* **351**, aad2622 (2016).
2. Ellis, E. C. *Ecol. Monogr.* **85**, 287–331 (2015).
3. Boivin, N. L. *et al. Proc. Natl Acad. Sci. USA* **113**, 6388–6396 (2016).
4. Ruddiman, W. F., Ellis, E. C., Kaplan, J. O. & Fuller, D. Q. *Science* **348**, 38–39 (2015).
5. Lewis, S. L. & Maslin, M. A. *Nature* **519**, 171–180 (2015).
6. Smith, B. D. & Zeder, M. A. *Anthropocene* **4**, 8–13 (2013).
7. Malm, A. & Hornborg, A. *Anthropocene Rev.* **1**, 62–69 (2014).
8. Bauer, A. M. & Bhan, M. *South Atl. Q.* **115**, 61–87 (2016).
9. Barry, A. & Maslin, M. *Geo Geog. Environ.* **3**, e00022 (2016).
10. Edgeworth, M. *Geol. Soc. Lond. Spec. Publ.* **395**, 91–108 (2014).



TALES OF WONDER

A bedtime story can ignite a lifelong love of science. *Nature* editors riffle through shelves and memories for favourites old and new.



George's Secret Key to the Universe

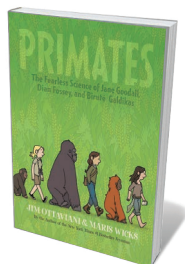
LUCY HAWKING & STEPHEN HAWKING
Doubleday: 2007.

Among the books that my family and I read on holiday, *George's Secret Key to the Universe* was the clear favourite with my daughters, aged five and eight. Theoretical physicist Stephen Hawking and his daughter, writer Lucy Hawking, have crafted a cracking narrative about a boy called George and his physicist neighbour, who owns a powerful computer named Cosmos. The story grabbed their attention — and held it for the week it took me to read it to them. They even preferred it to swimming. I was stunned. I hadn't ever tried a physics book with them before.

George's adventures with Cosmos, which can open a portal to any part of the Universe, are engaging thanks to humorous illustrations and high-resolution images of celestial objects including planets, the Milky Way and more. With George, we learn about the Solar System, the birth and death of stars, and black holes. Some facts are better than fiction. That is why my girls loved the story: it's about life, not just physics. George shares a world of discovery with us. And now my five-year-old wants to be an astronaut. My work is done.

May Chiao is the chief editor of *Nature Astronomy*.

ILLUSTRATIONS BY RICHARD WILKINSON



Primates: The Fearless Science of Jane Goodall, Dian Fossey, and Biruté Galdikas

JIM OTTAVIANI & MARIS WICKS
First Second: 2013.

This spring, my eight-year-old had to write a biography for school. She wanted her subject to be female, living and accomplished. At our public library, we came across *Primates* by Jim Ottaviani and Maris Wicks, a book about the lives and discoveries of primatologists Jane Goodall (chimpanzees), Dian Fossey (mountain gorillas) and Biruté Galdikas (orangutans). Sophia selected Galdikas, whose story fascinated her; she was also keen to write about someone unknown to her classmates.

A graphic novel is perfect for this triple narrative: it succinctly conveys a great deal of information that would be challenging with text alone. And it shows the excitement of fieldwork without glossing over its challenges or the obstacles facing female scientists. Ottaviani's engaging storytelling and Wicks's captivating illustrations transport the reader to the jungles and mountains where the three women did their groundbreaking work. We highly recommend this book for the budding zoologist or curious young mind in your own life.

Sophia Sykes-Finkelstein is in the third grade. **Joshua Finkelstein** is a former senior editor at *Nature* and current deputy director of the Allen Discovery Center at Tufts University.



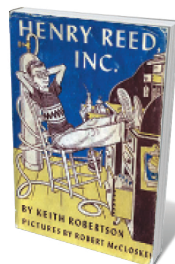
Insects Do the Strangest Things

LEONORA HORNBLow & ARTHUR HORNBLow
Random House Books for Young Readers: 1968.

It's the strangest things that stick in your head from childhood reading. For me, it was a series of books penned in the 1960s by Leonora and Arthur Hornblow, a New York couple who pushed kids to pause and ponder the living things around them by providing interesting facts to savour, remember and share. Did you know that some ants farm a fungus in their underground homes? Strange! And bee grubs will grow up to be queens if they are fed a steady diet of royal jelly. Each animal gets a brief, bite-sized chapter that covers a broad range of weird facts.

For me, *Insects Do the Strangest Things* was the most memorable book in the series. Unlike its companions on fish, birds and other animals, almost all of the creatures that it discusses were easy to find in my backyard. It was an entry point and an open invitation to dig around in the dirt and explore some more. And the warm, joyous illustrations by Michael K. Frith still remind me of summer nights listening to crickets, swatting at mosquitoes and wondering whether I ever really could, as the book suggests, catch enough fireflies to read it by.

Brendan Maher is a News Features editor at *Nature*.



Henry Reed, Inc.

KEITH ROBERTSON
Viking: 1958.

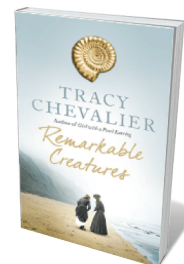
Keith Robertson's *Henry Reed, Inc.* (1958) was the first book that I read as a child in which the main character wanted to be a scientist when he grew up. (As *Nature* readers will know, such books remain vanishingly rare.)

Robertson's protagonist is the son of foreign-service parents, sent to spend a summer with his uncle just outside Princeton, New Jersey. Inspired by all the nearby research and development companies, such as Bell Laboratories, he founds his own research firm with Midge, a girl who lives across the street. Her father is a chemist — common in science fiction, but not children's literature.

The book is funny and touching (although, unfortunately, a bit late-fifties misogynistic) and the memory of it stuck. (I also strongly recommend the 1963 follow-up *Henry Reed's Journey*, which traces a pre-interstate-highway trip that Midge's family and Henry take from Los Angeles, California, to Princeton, after a conference attended by Midge's father.)

I read *Henry Reed, Inc.* to my son when he was about 6. Although he did not go on to become a scientist (that was crushed out of him in elementary school), he enjoyed it: it gave that sense of vistas of possibility.

Leslie Sage is a senior editor in physical sciences at *Nature* in Washington DC.



Remarkable Creatures

TRACY CHEVALIER
HarperCollins: 2009.

From her childhood, working-class woman Mary Anning hunted for fossils in England's Lyme Regis. She studied and sold the remains of large marine reptiles, many of them scientific firsts. *Remarkable Creatures*, Tracy Chevalier's novel inspired by Anning, shows how her finds drew the attention of renowned scientists such as geologist Charles Lyell and naturalist Georges Cuvier. Anning ultimately contributed to the theory of extinction, indicating its importance in Earth history.

We enjoyed searching for fossils in Spain's Aviaños mountains and were keen to learn about Anning. This stirring book shows that anybody can do scientific research, even against the odds. Science in the nineteenth century was dominated by men, and Chevalier shows how Anning and her self-educated friend Elizabeth Philpot battled for recognition. She communicates accurate science in a way sure to interest young adults who might shy away from academic papers. Chevalier takes liberties in describing the relationships between Anning, Philpot and other historical figures — but isn't this the storyteller's privilege? *Remarkable Creatures* made us yearn to visit Lyme Regis and hunt fossils again.

Laura Mata Le Bot (14) is a student with a love of science, dance and experimental baking. **Nathalie Le Bot** is a senior biology editor at *Nature*.



Outside Your Window: A First Book of Nature

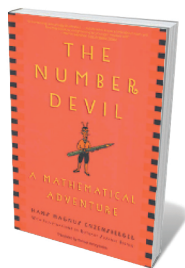
NICOLA DAVIES
Candlewick: 2012.

Ella, age four, loves poking around in mud and searching for minibeasts under stones. *A First Book of Nature* is her sort of book. Beautiful illustrations by Mark Hearld and words by Nicola Davies send the reader on a mesmerizing journey through the seasons, from melting icicles to the sounds of sleepy summer, through berry picking to slow wintertime.

Poetry peppers the colourful pages, perfectly capturing the wonder of discovery. The animals are carefully observed. Lizards are “fast as thinking/There, then not there”. A running horse is thrillingly scary, but at rest and offered a carrot, “its dark eye is quiet and its nose is velvet, more tender than your own cheek”.

The odd science lesson is thrown in — how rainbows appear, the life cycles of fungi. There are suggestions for activities (“Things to do in your den. 1. Sit and think. 2. Notice things, like the smell of the earth, or what beetles are doing.”) There are even recipes for bird cake and compost. The book is a work of art that holds Ella’s rapt attention throughout. And it reminds me that even on a grey London day, I should take a break and see what’s outside.

Clare Thomas is a senior editor in biology at Nature.



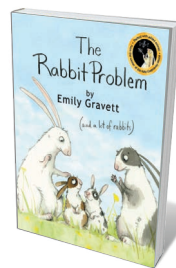
The Number Devil: A Mathematical Adventure

HANS MAGNUS
ENZENSBERGER
Henry Holt: 1997.

At school, my best friend’s favourite book was Hans Magnus Enzensberger’s *The Number Devil*. Her enthusiasm made me curious enough to give the story a try. In the book, a charismatic devil called Teplo-taxl visits the dreams of 12-year-old Robert to introduce him to the magical world of mathematics. From prime numbers (which the devil calls “prima donnas”) and vroom numbers (factorials, in mundane English) to the concept of mathematical proofs, Enzensberger’s descriptions were brilliantly imaginative. I, too, soon found myself thinking about the weird connections between triangular arrays and Fibonacci numbers whenever a maths lesson at school became dull.

Fast-forward 15-odd years and my friend, having earned a maths PhD, has just given up a lucrative consulting job to work in the didactics of mathematics. If you ask me, there is at least a subconscious connection between that decision and Enzensberger’s book. Explaining the mind-boggling abstractness of her algebraic topology research with the most colourful images, my friend has, in any case, made me firmly believe that everybody can understand mathematics — as long as they have access to a number devil.

Leonie Mueck is a senior editor in physical sciences at Nature.



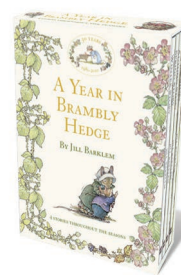
The Rabbit Problem

EMILY GRAVETT
Macmillan Children’s Books: 2009.

Emily Gravett is a master of delicate, humorous water-colour illustration, and her trademark skills (also on display in 2011’s *The Odd Egg* and the 2008 *Spells*, both published by Macmillan Children’s Books) are well to the fore in *The Rabbit Problem*. This faux calendar illustrates Fibonacci’s sequence, using the same example that Fibonacci himself employed in his classic of arithmetic, the 1202 *Liber abaci* (*Book of Calculation*). In Fibonacci’s theorem, the next number in the sequence is generated by summing the two previous numbers (1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144 ...). My eight-year-old could easily handle the maths.

Starting with one pair of rabbits in a field, Gravett demonstrates how many of the animals there will be after a year (“No Rabbits May Leave The Field”). The book addresses population problems month by month, from Hungry Rabbits (solution: plant carrot seed) to Cold Rabbits (knit a jumper). It all ends with December’s 144 pairs dramatically escaping in a double-page pop-up, pop-out spread. The charming inserts such as (partly chewed) carrot recipes and leporine newspaper articles make it suitable as a simple bunny tale for tots, too.

Dinah Loon is a physical-sciences subeditor at Nature.



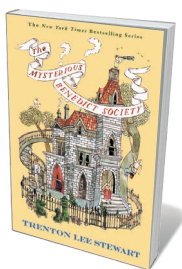
A Year in Brambly Hedge

JILL BARKLEM
HarperCollins: 2010.

Jill Barklem’s Brambly Hedge books follow a community of mice through Britain’s changing seasons. From the Snow Ball of *Winter Story* to the surprise picnic of *Spring Story* and riverside wedding of *Summer Story*, our diminutive hedge-dwellers mark the passing of the year in parties wreathed in buds, blossoms, berries and ice. Throughout, they pick and store nature’s bounty, turning it into miniature pies and pastries, jellies and syllabubs. Small wonder that the younger of my own small wonders (Josh) demanded a Brambly Hedge party for his fifth birthday, complete with a rowan-trimmed straw hat.

Somehow, Barklem’s quartet of stories — although ostensibly about the festivities of bonnet-clad rodents — dodges tweeness and lands somewhere deeper and more resonant. Her exquisitely illustrated books are as much hymns to the joys of observing the natural world, to communing with it and in it, as are the works of travel writer Robert MacFarlane or environmental historian Oliver Rackham. My wee boy loves to point out how this tree in the story we’re reading has fruit, whereas in the previous book it was in flower, or to assure me that we, too, could nap beneath bluebells. Quite so.

Sara Abdulla is chief commissioning editor for Nature.



The Mysterious Benedict Society

TRENTON LEE STEWART
Little, Brown (Books for Young Readers): 2007.

Reynie Muldoon is a 12-year-old orphan who replies to a newspaper advertisement inviting children to attend a series of mind-bending tests. He passes and, with three other gifted children, enters the service of the kindly yet eccentric Nicholas Benedict, who is investigating the nefarious doings of the Learning Institute for the Very Enlightened (L.I.V.E.), a school run by Benedict's evil twin, Ledropha Curtain. Reynie and his companions have to infiltrate the school and find out what's going on.

One could dismiss *The Mysterious Benedict Society* as espionage for pre-teens — an urban version of Arthur Ransome's *Swallows and Amazons* (1930) — except for the utter eccentricity of all the characters, and the full-fat science and logic content. (In one test, the children are asked to navigate a tiled floor without treading on any of the squares. Only Reynie walks blithely across — the tiles are rectangular.)

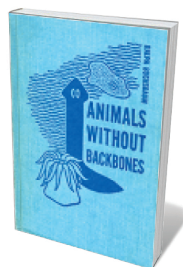
Like all the best books for children, it's done the old-fashioned way: without compromise. This volume is the first in what is currently a four-part series — one instalment of which is called (and I'm not making this up) *The Mysterious Benedict Society and the Prisoner's Dilemma*.

Henry Gee is senior editor in biology at Nature.



Animals Without Backbones: An Introduction to the Invertebrates

RALPH BUCHSBAUM
University of Chicago Press: 1938.



If you had told the ten-year-old me that *Animals Without Backbones* was a classic biology textbook, I would never have picked it up. I struggled to, anyway: the decades-old dog-eared copy that I found at the back of my dad's bookcase had a loose monochrome cover that always wanted to come off in my hands. This was a link to his world as a scientist, and to what he did all day. (Actually, he was a research chemist, but what did I know.)

More, the book was a glimpse of a world just as alien as those in the pages of my *2000 AD* comic, peopled with warlocks and genetic infantrymen. The pictures looked hand-drawn, and showed features on the outside of the creatures as well as their inner structures. I studied those pages and copied the drawings — the stunning representation of the *Hydra* especially — into my sketch pad, next to *Rogue Trooper* and *Judge Dredd*. *2000 AD* later published one of those drawings, but it was the fantastic stories of the true, hidden world of invertebrates that really fired my imagination.

David Adam is Editorials editor at Nature.



Doctor Dolittle in the Moon

HUGH LOFTING
Jonathan Cape: 1928.

Growing up in wild corners of North America, I had close encounters with bears, deer and rattlesnakes — as well as a huge range of fictitious fauna in books from Beatrix Potter's 1904 *The Tale of Two Bad Mice* to Rudyard Kipling's 1902 *The Cat that Walked by Himself*. Hugh Lofting's *Doctor Dolittle* series wonderfully mixed it all up, harnessing the empirical method to unlock an alternate universe of beasts that behave like beasts, but engage in rational discourse. A favourite was the 1928 *Doctor Dolittle in the Moon*. In it, our interspecies polyglot shifts his research into a higher gear.

Dolittle's lunar mission is a naturalist's dream, with a gargantuan moth and vast oxygen-emitting flowers serving as spacecraft and life support, respectively. With a crack team from his household in Puddleby-on-the-Marsh, the doctor roams the Moon digging up data on its formation, the impacts of microgravity and the longevity of its exotic, surprisingly numerous inhabitants. Perhaps most delightful is Dolittle's encounters with the Whispering Vines — voluble vegetation living in ecological cooperation. Although I didn't know it then, Lofting's free-floating imagination had hit upon some of today's more fascinating questions in plant science.

Barbara Kiser is Nature's Books and Arts editor.

Correspondence

Urgent action on Cerrado extinctions

Brazil has reduced deforestation in the Amazon by almost 80%, a major contribution to the goals of the United Nations Convention on Biological Diversity (CBD). Neighbouring Cerrado, a vast tropical savannah that hosts some 4,800 species of plants and vertebrates found nowhere else, has not fared so well.

Our findings show that a severe extinction episode is unfolding in the Cerrado, with plant extinctions projected to be an order of magnitude higher than all global recorded plant extinctions so far — yet in our view, this catastrophe can be avoided (B. B. N. Strassburg *et al.* *Nature Ecol. Evol.*, in the press).

Public and multi-stakeholder policies that supported Brazil's Amazon success need to be replicated in the Cerrado. For example, the region could be included in the Amazon's soya-bean moratorium and its public protected areas increased from 7.5% of the biome to the nationwide CBD target of 17%. Also, the new Forest Code should be enforced there, and large-scale restoration undertaken to boost and connect habitats that are crucial for endangered species.

These and other policies could transform an impending irreversible extinction episode of global proportions into another immense contribution to national and international goals of safeguarding biodiversity.

Bernardo B. N. Strassburg, *Agnieszka Latawiec International Institute for Sustainability; and Pontifical Catholic University, Rio de Janeiro, Brazil.*

Andrew Balmford *University of Cambridge, UK.*
b.strassburg@iis-rio.org

Fermi's predictions live on

Gino Segrè and Bettina Hoerlin's biography of Enrico Fermi — *The Pope of Physics* (Henry Holt,

2016) — is a delightful sequel to the inspiring *Atoms in the Family*, written by Fermi's wife Laura in 1954 (Univ. of Chicago Press). I find reviewer Graham Farmelo's judgement too harsh on the fallibility of Fermi's foresight (see *Nature* **538**, 168–169; 2016).

In 1934, Fermi predicted (in a paper rejected by *Nature*) that particles actually change their identity in weak nuclear interactions; with his student C. N. Yang in 1949, he suggested that there were too many particles for them all to be elementary. These remarkable testaments to Fermi's foresight have survived as founding concepts of today's standard model of matter.

Twenty years later, I generalized the Fermi–Yang model by assuming that quarks combine to form neutral composite states that interact only over short distances (P. C. M. Yock *Int. J. Theor. Phys.* **2**, 247–254; 1969). A similar solution was later adopted in the theory of coloured quarks (H. Fritzsch *et al. Phys. Lett.* **47B**, 365–368; 1973) on which the standard model is based.

Philip Yock *Auckland, New Zealand.*
p.yock@xtra.co.nz

Forge a clearer path for technical careers

The UK Biotechnology and Biological Sciences Research Council (BBSRC) surveyed more than 800 technicians and laboratory assistants about their roles, careers and professional recognition (see go.nature.com/2fs6icb). The results reveal opportunities for change in the current ill-defined status of technical staff.

These non-faculty members professed to a blurring between technical and academic roles. Forty per cent of them teach and 60% supervise students. They need to have expertise in mathematics, statistics and computing, for example — skills that are crucial for research

but hard to come by (see go.nature.com/2grqk10). Yet the progression opportunities and career pathways for technicians and academics are poles apart.

Academic career metrics such as publications and grants are not necessarily useful for technical staff. We found that 80% of those surveyed had contributed to research papers, although only 19% were first authors; 22% had written grants, but only 12% of respondents realized that they could apply for BBSRC funding. Moreover, 39% did not have an up-to-date job description and just 32% said that their role carried clear performance-related metrics.

Technical staff should not remain the unsung heroes of research. They need a well-defined career structure, clear job descriptions, appropriate performance metrics and an associated reward system. The BBSRC will seek to develop sustainable careers, in partnership with research organizations and according to the professional registration standards of the Science Council (www.sciencecouncil.org).

Michael Ball, Rob Hardwick *BBSRC, Swindon, UK.*
Kelly Vere *Science Council, London, UK.*
michael.ball@bbsrc.ac.uk

Nobel launchpad is immaterial

The question of which institutions produce the most Nobel prizewinners is one not just of numbers but of ratios (see *Nature* **538**, 152; 2016). Merely counting doesn't cut it.

We analysed the distribution of US Nobel laureates and members of the US National Academy of Sciences according to their undergraduate institutions (R. Root-Bernstein and K. Pawelec *J. Genius Eminence* **1**, 28–42; 2016). We found that most had trained in California, Illinois, Wisconsin, Massachusetts, New York,

Connecticut and New Jersey. Yet after correcting either for the number of scientists per state or by state population, the probability of becoming a Nobel prizewinner or National Academy member was almost the same, irrespective of institution.

The case is entirely different for graduate education, where institution seems to be a dominant factor in future success.

Robert Root-Bernstein *Michigan State University, East Lansing, USA.*
rootbern@msu.edu

Focus like a lens, not like a laser

As someone who once held the world's first laser in my hand (while interviewing its inventor, the late Ted Maiman), I find the expression 'laser-focused' irritating. That phrase, or the equally misleading 'focused like a laser', appears at least twice in *Nature*'s 27 October 2016 issue.

The *Oxford English Dictionary* defines 'focus' as "to converge to or towards a single point or place", which a laser does not. It produces coherent light — light of the same wavelength and in phase; no converging is involved. To speak of a laser focusing light is to mislead generations of schoolchildren (and the rest of us) into thinking that a laser works like a lens, missing the beautiful quantum process that is actually involved.

Maiman's laser was simply a cylinder of artificial ruby about the size of a cigarette filter, half-silvered on both ends and nested in the spiral of a high-intensity flashbulb, all contained in an aluminium can no bigger than a flashlight battery. *Nature* had the good sense to publish his paper on the invention (*Nature* **187**, 493–494; 1960) after another journal had rejected it.

Richard Rhodes *Half Moon Bay, California, USA.*
richardrhodes1@comcast.net

FORUM: Climate science

The history of Greenland's ice

Global sea levels would rise by several metres if the Greenland Ice Sheet melted completely. Two studies have examined its past behaviour in an effort to evaluate its vulnerability in a warming world — and have come to seemingly conflicting conclusions. Two geochemists and a glaciologist discuss the issues. [SEE LETTERS P.252 & P.256](#)

THE PAPERS IN BRIEF

- Knowledge of the ancient history of ice sheets is needed to inform predictions of their future response to climate change.
- On page 252, Schaefer *et al.*¹ present measurements of beryllium-10 (^{10}Be) and aluminium-26 (^{26}Al) from bedrock beneath the Greenland Ice Sheet.
- From their data, they propose that Greenland has undergone one or more

episodes of deglaciation during the past 1.1 million years.

- On page 256, Bierman *et al.*² present ^{10}Be and ^{26}Al data from marine sediment cores collected off the coast of Greenland.

- They conclude that the East Greenland Ice Sheet remained present during the Pleistocene (the epoch that lasted from 2.6 million to 11,700 years ago), but grew and shrank dynamically in response to climate.

increase in the volume of the GrIS.

In parallel, Schaefer *et al.* analysed bedrock samples recovered from beneath several kilometres of ice in the centre of the present-day GrIS. These rocks cannot have been transported to their current position from elsewhere. The presence of ^{10}Be and ^{26}Al in these samples is therefore direct evidence for past deglaciation episodes at this location.

Schaefer and colleagues used the observed concentrations and ratios of ^{10}Be and ^{26}Al in a numerical model to explore plausible glaciation–deglaciation scenarios. They firmly conclude that the sampled bedrock must have been ice-free for at least one episode of about 280,000 years during the past 1.1 million years. Other glaciation–deglaciation scenarios are more plausible, however, such as several shorter disappearances (several thousand years) of the GrIS during the warmest interglacial periods of the Pleistocene³.

The deglaciation proposed by Schaefer

Cosmic signature

PIERRE-HENRI BLARD & GUILLAUME LEDUC

It is difficult to find evidence for past waxing and waning of the Greenland Ice Sheet (GrIS), because the glacial debris that can provide a record of ice-sheet dynamics (Fig. 1) is erased each time glaciers advance after a period of deglaciation. Schaefer *et al.* and Bierman *et al.* now report the presence of two rare nuclides (^{10}Be and ^{26}Al) in Greenland's geological archives. These 'cosmogenic' nuclides are produced by the interaction between atoms in minerals and high-energy cosmic neutrons that probably originated in supernovae, and they provide much-needed evidence about the GrIS's ancient past.

A key characteristic of these isotopes is that they are produced in detectable quantities in only the first few metres below Earth's surface. If they become buried below hundreds of metres of ice, their concentrations decrease through radioactive decay. The half-lives of ^{10}Be and ^{26}Al are 1.4 million and 0.7 million years, respectively, so the presence of these nuclides in bedrock under ice sheets is direct evidence that one or several ice-free episodes occurred in the recent past (less than 5 million years ago).

Bierman *et al.* report variations in ^{10}Be and ^{26}Al from marine sediments deposited near the eastern margin of today's GrIS over the past 7.5 million years. The low ^{10}Be concentrations suggest that the sedimentary material originates from locations that were efficiently

shielded from cosmic rays. The authors therefore conclude that large ice caps probably existed continuously across East Greenland — the source of the sedimentary material — during the past 7.5 million years. They also observe an additional long-term decrease in ^{10}Be and ^{26}Al concentrations during this period, which indicates a gradual



NICK COBBING/GREENPEACE

Figure 1 | Ice-sheet dynamics. These icebergs calved from the Helheim Glacier, an outlet glacier at the periphery of the Greenland Ice Sheet. Such glaciers are responding rapidly to climate change.

et al. seems at odds with Bierman and colleagues' conclusions. But small ice caps might have persisted on the East Greenland highlands during such deglaciated episodes, and could be the origin of the ¹⁰Be-poor material observed in Bierman and co-workers' marine sediments. Indeed, modelling experiments⁴ estimate that ice caps 1,000 metres thick might persist on the south-eastern Greenland highlands in the future even if 95% of the GrIS disappears (Fig. 2).

Could the GrIS undergo similar deglaciations in the future? Average global warming of just 2 °C compared with preindustrial temperatures could be enough to melt more than 75% of the GrIS³. If this high sensitivity to warming is confirmed, then the current worst-case scenarios⁵ for future sea-level rise associated with anthropogenic global warming will need to be revised.

Pierre-Henri Blard is at the Centre de Recherches Pétrographiques et Géochimiques, CNRS, Université de Lorraine, 54500 Vandœuvre-lès-Nancy, France. **Guillaume Leduc** is at the Centre Européen de Recherche et d'Enseignement des Géosciences de l'Environnement, (Aix-Marseille University, CNRS, IRD), Europôle Méditerranéen de l'Arbois, 13545 Aix en Provence, France. e-mails: blard@crpg.cnrs-nancy.fr; leduc@cerege.fr

Glaciological puzzle

NEIL GLASSER

One of the most powerful methods for predicting future behaviour is to look for clues from the past. But interpreting the past is rarely straightforward, as illustrated by the seemingly contradictory conclusions of Schaefer *et al.* and Bierman and colleagues. The findings can perhaps be reconciled, but that would require glaciologists to have a major rethink about current theories of ice-sheet behaviour.

Schaefer and colleagues' analysis of the terrestrial bedrock beneath the GrIS leads them to conclude that the ice sheet was sometimes reduced to less than 10% of its current volume, so that the only ice to be found in Greenland was in the peripheral mountain

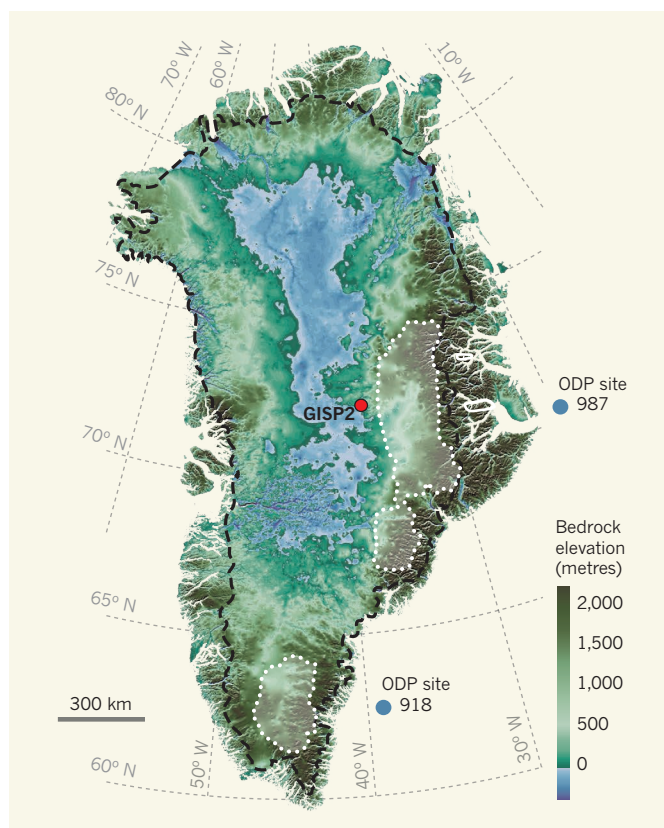


Figure 2 | Subglacial topography of Greenland and deglaciation of its ice sheet. The map shows the bedrock core of Greenland and its elevation above or below sea level. Schaefer *et al.*¹ analysed bedrock samples taken from a site at Greenland's centre (GISP2), and conclude that Greenland was ice-free for long periods during the Pleistocene epoch (2.6 million to 11,700 years ago). Bierman *et al.*² analysed marine sedimentary cores taken from two sites off the east coast of Greenland (ODP sites 918 and 987), and deduce that the East Greenland Ice Sheet remained intact during the Pleistocene. Although seemingly contradictory, the results could be reconciled if a small ice cap remained during major deglaciations — previous simulations⁴ suggest that remnants of the ice sheet (enclosed by dotted white lines) would persist in the eastern highlands even if 95% of the Greenland Ice Sheet melted. The broken black line indicates the approximate extent of the current ice sheet. (Adapted from refs 4 and 7.)

ranges. By contrast, Bierman and co-workers' study of the marine record indicates that warming was short-lived, and that it seldom caused substantial and lengthy reductions in the extent of the East Greenland Ice Sheet. So how do these conclusions fit in with what is known about ice-sheet dynamics?

One source of information is numerical modelling of ice sheets. Simulations⁶ based on such models tend to predict that the GrIS existed continuously throughout the Pleistocene. The models rarely predict extended periods during which Greenland was nearly ice-free.

Glaciologists also know from observations that the GrIS is dynamic, and that its peripheral outlet glaciers are responding rapidly to climate change (Fig. 1). But most of these observations come from areas around the periphery of the ice sheet; the central areas are considered to be relatively stable. The

conclusion that the GrIS shrank to less than 10% of its current volume, leaving ice-free areas at its centre, is therefore particularly difficult to explain glaciologically because it requires the entire ice sheet to have been removed and completely 'rebuilt', possibly on several occasions.

Perhaps part of the explanation lies in Greenland's subglacial topography⁷. Earth's crust is depressed at the centre of Greenland because of the great weight of the ice sheet, whereas the periphery is surrounded by mountain ranges (Fig. 2). Deglaciation right to the centre of the GrIS — rather than just changes in the peripheral outlet glaciers — would therefore require massive, rapid changes of ice dynamics. This calls for glaciologists to rethink how the ice sheet interacts with subglacial topography and sea level. Could the same deep valleys that give the central GrIS its great thickness also open up the underside of the ice sheet and make it vulnerable to warming oceans as sea levels rise?

These new papers throw down three immediate challenges. First, we must seek ways to reconcile the two seemingly contradictory records of the ice sheet's past behaviour. Second, we must try to understand the dynamical processes of the ice sheet that make possible the required huge and rapid variations in the size and volume of the GrIS. And third, we need to assess whether such variations could happen again in the near future, with all the

attendant social and economic consequences that would accompany a rapid rise in global sea level. ■

Neil Glasser is in the Department of Geography and Earth Sciences, Aberystwyth University, Aberystwyth SY23 3DB, UK. e-mail: nfg@aber.ac.uk

1. Schaefer, J. M. *et al.* *Nature* **540**, 252–255 (2016).
2. Bierman, P. R., Shakun, J. D., Corbett, L. B., Zimmerman, S. R. & Rood, D. H. *Nature* **540**, 256–260 (2016).
3. Dutton, A. *et al.* *Science* **349**, aaa4019 (2015).
4. Fyke, J., Eby, M., Mackintosh, A. & Weaver, A. *Clim. Dyn.* **43**, 2249–2260 (2014).
5. Ridley, J., Gregory, J. M., Huybrechts, P. & Lowe, J. *Clim. Dyn.* **35**, 1049–1057 (2010).
6. Calov, R., Robinson, A., Perrette, M. & Ganopolski, A. *Cryosphere* **9**, 179–196 (2015).
7. Morlighem, M., Rignot, E., Mouginot, J., Seroussi, H. & Larour, E. *Nature Geosci.* **7**, 418–422 (2014).

A parasite's parasite saves host's neighbours

Viruses can be attacked by parasitic viruses, which compete with them for cellular resources. It emerges that one such parasitic virus can defend a host-cell population from a viral attack. [SEE LETTER P.288](#)

EUGENE V. KOONIN & MART KRUPOVIC

So nat'ralists observe, a flea
Has smaller fleas that on him prey;
And these have smaller fleas to bite 'em.
And so proceeds *ad infinitum*.

These words, written by the satirist Jonathan Swift in *On Poetry: A Rhapsody*¹, describe a real biological phenomenon whereby a parasite is itself subject to parasitic attack². The interactions that occur when a parasite is preyed on are poorly understood — yet anything that limits the success of the primary parasite might affect the outcome of infection. On page 288, Fischer and Hackl³ provide insight into the relationship between a host cell, a virus and a parasitic virus that preys on the virus.

Single-celled organisms called protists can be infected by giant viruses of the family *Mimiviridae*⁴. Giant viruses reproduce in a membrane-bound structure known as the virus factory, which they establish in the cytoplasm of their protist host. Giant viruses are themselves preyed on by parasitic viruses called virophages⁵, from the family *Lavidaviridae*⁶. Virophages also reproduce in the virus factory, and substantially reduce the reproduction rate of the giant viruses⁵.

Fischer and Hackl investigated the infection of the marine protist *Cafeteria roenbergensis* with the *C. roenbergensis* giant virus CroV and its associated virophage, mavirus. Mavirus can enter the host cell independently of CroV uptake⁷. The authors co-infected a cell population with CroV and mavirus, and found that the mavirus genome was inserted into the

protist genome in approximately one-third of the cells, providing the first direct evidence of virophage DNA integration into the genome of a cellular host. The *C. roenbergensis* strain analysed carried 11 copies of the mavirus genome, integrated at different chromosomal locations. Mavirus and other virophages encode enzymes known as integrases that enable DNA insertion into a host genome^{7,8}.

Does virophage DNA that has been integrated into the protist genome have any function in the host cell? Fischer and Hackl observed that when protist cells containing integrated mavirus genome were infected with CroV, the mavirus genes were expressed and mavirus replicated to form new viral particles, resulting in reactivation of mavirus in the cell. The authors propose that a CroV-encoded transcription factor is responsible for mavirus reactivation, consistent with a previous observation⁷ that CroV and mavirus have similar promoter sequences that drive gene expression.

Fischer and Hackl found that the reactivation of mavirus that followed CroV infection did not prevent the replication of CroV, and host cells infected with CroV still died, releasing virus particles of both CroV and mavirus into the environment. Presumably, reactivation of the integrated mavirus occurs late in the reproductive cycle of CroV and therefore cannot block CroV reproduction. However, the released mavirus decreased CroV reproduction in a subsequent round of infection, thereby preventing the spread of CroV in the protist population. Although a cell infected by CroV seems to be doomed, irrespective of the presence of mavirus, the death of this cell can protect related neighbouring cells from destruction by the giant virus (Fig. 1).

What kills host cells co-infected with CroV and mavirus is unknown. Is it CroV, despite its reproduction being limited by the presence of mavirus in the virus factory, or is it mavirus itself? Reactivation of mavirus in response to CroV infection seems to represent a case of an altruistic defence mechanism of the host, in which a cell dies, releasing mavirus that can then protect the neighbouring population of related cells. The existence and role of altruistic defence mechanisms in such a context remain controversial, given the continued debate about the importance of kin selection in evolution, especially in unicellular organisms⁹. Nevertheless, potential examples in this category are increasing, and include the activation of programmed cell death by infection¹⁰.

Mavirus protection of the *C. roenbergensis* population against CroV spread can also be regarded as a case of adaptive immunity, which involves immunological memory of past infections. In microbes' CRISPR–Cas-mediated adaptive immune system¹¹, viral DNA sequences encountered

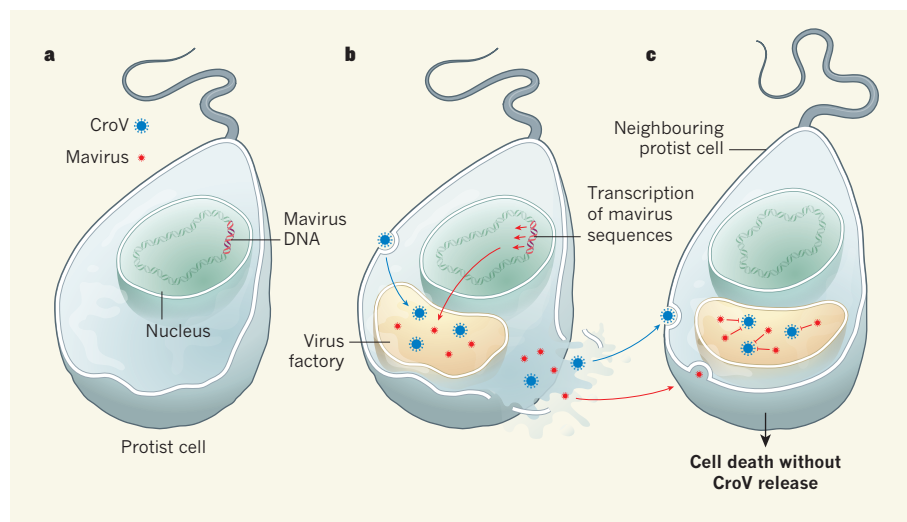


Figure 1 | A parasitic virus can protect a host-cell population from viral attack. Fischer and Hackl³ investigated the relationship between the unicellular protist *Cafeteria roenbergensis*, the *C. roenbergensis* giant virus CroV, which infects this host, and mavirus, a parasitic virus that targets CroV by suppressing its replication. **a**, When the authors infected *C. roenbergensis* with mavirus and CroV, they observed integration of the mavirus genome into the genome of about one-third of the host cells. **b**, When cells with integrated mavirus were subsequently infected with CroV, mavirus sequences were transcribed. Mavirus particles then formed and replicated in a membrane-bound structure in the cytoplasm known as the virus factory, where CroV also replicates, but did not block the replication of CroV. The cell eventually broke open and released virus particles into the environment. **c**, The released viruses were taken up by neighbouring cells. Although nearly every cell infected with CroV seems to die, Fischer and Hackl observed increased survival of the host-cell population in these subsequent cellular infections, presumably because mavirus suppresses CroV reproduction in the virus factory. Thus, mavirus can protect host-cell populations from CroV attack, in a similar way to how the memory of a previous attack can provide protection in acquired immunity, as occurs in microbial CRISPR–Cas systems.

during an attack are retained by cells and used to prevent subsequent attacks associated with the same or closely related DNA sequences. The twist involved when mavirus is the memory aid is that the infectious agent is remembered indirectly, by host integration of the virophage DNA sequences, which are expressed only during subsequent encounters with the giant virus.

As with CRISPR–Cas-mediated immunity, it is not clear how a host cell survives to retain immunological memory, given that infection by CroV is usually fatal, as Fischer and Hackl demonstrate. In the CRISPR–Cas system, immunological memory is thought to be formed when cells are infected with defective phage particles¹², and such abortive infection might also occur in the CroV–mavirus system. Another possibility is that mavirus genome integration into the host genome can occur in the absence of CroV co-infection.

Many questions remain. Perhaps the most pressing one is: how general is this antiviral defence mechanism discovered by Fischer and Hackl? The genome of the unicellular green alga *Bigeloviella natans* contains numerous integrated virophage genomes¹³, and this possibly provides another example of a virophage-mediated host defence. The current observations might be only the tip of the iceberg.

The mavirus virophage belongs to a large class of Polinton-like viruses and mobile genetic elements that can integrate into the genomes of diverse organisms^{8,14}. Do some — or even most — of this class that have not, so far, been observed to produce virus particles provide defence against viral infections? The integration of viral DNA sequences, known as endogenous viral elements, is a common outcome when viruses with RNA or DNA genomes infect cells¹⁵. Were these endogenous viral elements obtained fortuitously as a result of aberrant genome integration, or might they have been retained by a host to provide immunological memory? The experiments needed to address these intriguing questions have not yet begun in earnest.

The relationship between *C. roenbergensis*, CroV and mavirus highlights two fundamental biological principles: the pervasiveness of parasites that have evolved to target almost all replicating biological entities, and the use of altruistic modes of protection that are probably intrinsic to cellular life. There is still much to learn about the mechanisms of attack and defence that operate in these fascinating and complex struggles for survival. ■

Eugene V. Koonin is at the National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA. **Mart Krupovic** is in the Department of Microbiology, Institut Pasteur, Paris 75015, France.

e-mail: koonin@ncbi.nlm.nih.gov

1. Swift, J. *On Poetry: A Rhapsody* (Huggonson, 1733).
2. Parratt, S. R. & Laine, A.-L. *ISME J.* **10**, 1815–1822 (2016).
3. Fischer, M. G. & Hackl, T. *Nature* **540**, 288–291 (2016).
4. Fischer, M. G. *Curr. Opin. Microbiol.* **31**, 50–57 (2016).
5. Desnues, C., Boyer, M. & Raoult, D. *Adv. Virus Res.* **82**, 63–89 (2012).
6. Krupovic, M., Kuhn, J. H. & Fischer, M. G. *Arch. Virol.* **161**, 233–247 (2016).
7. Fischer, M. G. & Suttle, C. A. *Science* **332**, 231–234 (2011).
8. Yutin, N., Shevchenko, S., Kapitonov, V., Krupovic, M.

- & Koonin, E. V. *BMC Biol.* **13**, 95 (2015).
9. Allen, B., Nowak, M. A. & Wilson, E. O. *Proc. Natl Acad. Sci. USA* **110**, 20135–20139 (2013).
 10. Durand, P. M., Sym, S. & Michod, R. E. *Curr. Biol.* **26**, R587–R593 (2016).
 11. Mohanraju, P. *et al. Science* **353**, aad5147 (2016).
 12. Hynes, A. P., Villion, M. & Moineau, S. *Nature Commun.* **5**, 4399 (2014).
 13. Blanc, G., Gallot-Lavallée, L. & Maumus, F. *Proc. Natl Acad. Sci. USA* **112**, E5318–E5326 (2015).
 14. Krupovic, M. & Koonin, E. V. *Nature Rev. Microbiol.* **13**, 105–115 (2015).
 15. Katzourakis, A. & Aswad, A. *Curr. Biol.* **26**, R427–R429 (2016).

ASTROPHYSICS

Elemental abundances across cosmic time

The chemical composition of a massive galaxy in the early Universe reveals an extremely short period of star formation. This result could challenge our ideas about the evolution of galaxies and of the Universe itself. SEE LETTER P.248

CHIYAKI KOBAYASHI

Stars are fossils that retain the history of their host galaxies. At the end of their lives, they explode as supernovae, producing heavy elements that are distributed into the surrounding interstellar gas. New stars that are created from this gas contain the elements that were produced from the previous generations of stars. By analysing the abundance patterns of the elements, it is therefore possible to determine how many and what kind of supernovae exploded in the past. On page 248, Kriek *et al.*¹ estimate elemental abundances of a massive galaxy, as seen 11 billion years ago, which suggest that this galaxy formed by a short and intense burst of star formation and then suddenly ‘died’ without producing more stars. The authors’ findings could reshape our theories of galaxy formation².

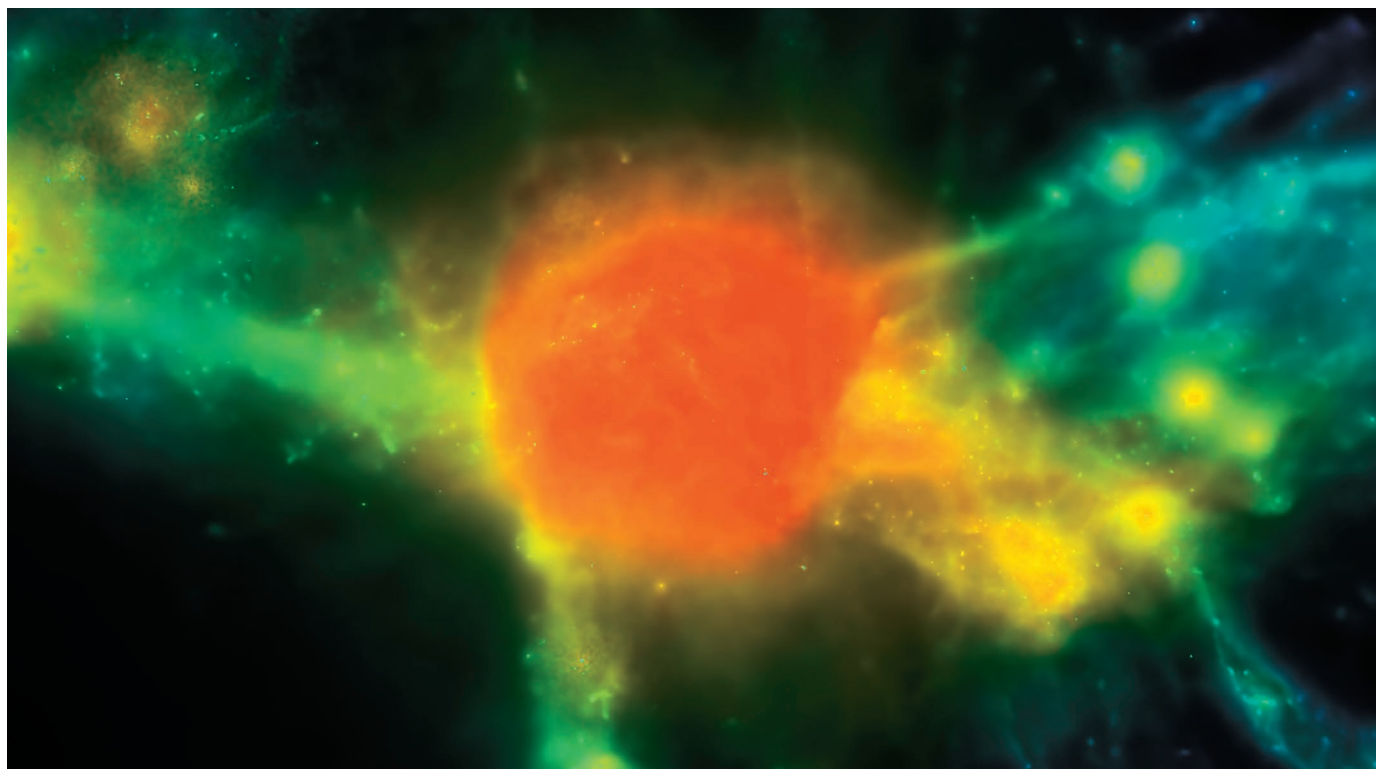
Explaining the origin of the elements is one of the scientific triumphs at the interface of nuclear physics and astrophysics. As the astronomer Fred Hoyle predicted³, carbon and heavier elements were not produced during the Big Bang, but are instead created inside stars. The α -elements — oxygen, magnesium, silicon, sulfur and calcium — are produced by massive stars (those approximately ten times more massive than the Sun) before being ejected in supernovae. Conversely, iron (Fe) is mainly produced by a different type of stellar explosion called a type Ia supernova. Theoretical models can successfully reproduce the observed abundances of these elements in nearby stars of the Milky Way⁴.

Elements heavier than zinc can form by neutron-capture processes in stars, or by other exotic astronomical events⁵.

This knowledge of nuclear astrophysics has been used to determine the formation and evolutionary histories of nearby (low redshift) galaxies, using an approach called galactic archaeology. Kriek and colleagues show that this approach can also be applied to distant (high redshift) galaxies. The authors study a massive galaxy that has a relatively large redshift of 2.1 and find a surprisingly high ratio of α -elements to Fe. This result suggests that the galaxy had an intense period of star formation that lasted for only between 0.1 billion and 0.5 billion years (approximately 10% of the age of the Universe at that time) before type Ia supernovae began to occur.

The nature of the physical mechanism that causes such quenching of star formation has been one of the big questions in astronomy for more than 20 years. One suggested mechanism is feedback from the supermassive black holes at the centres of massive galaxies (Fig. 1) — observed as active galactic nuclei (AGN). AGN-driven winds would have removed the surrounding interstellar gas, preventing new stars from forming. However, the high α /Fe ratio in massive galaxies has not yet been fully explained by numerical simulations of galaxies that include this feedback mechanism⁶.

Kriek and collaborators’ results make the situation even more complicated. Their galaxy is observed as it was when the Universe was only 3 billion years old. At this time, black holes might not have been large enough to produce sufficient feedback to prevent star



PHILIP TAYLOR AND CHIAKI KOBAYASHI

Figure 1 | Star formation cut short. Kriek *et al.*¹ have studied a massive galaxy in the early Universe that formed stars over a relatively short period. The quenching of star formation that occurs after this period could have been caused by feedback from the supermassive black hole at the centre of the galaxy, which would have expelled the surrounding interstellar gas and prevented new stars from forming. Shown here is a simulation of such a galaxy, including the feedback effect (see text). The colours represent the temperature of gas from low (blue) to high (red).

formation, because black holes are expected to co-evolve with galaxies. This co-evolution is suggested by the tight correlation between the masses of galaxies and their black holes. It is also consistent with the space density of quasars — optically bright AGN — and cosmic star-formation rates, both of which decrease sharply⁷ beyond redshifts of 2, where the authors' galaxy is found.

Thanks to quantum mechanics and atomic spectroscopy, elemental abundances can be estimated from absorption lines in stellar spectra. However, it is not straightforward to estimate these abundances even in the spectrum of a single star because of complex fluid mechanics (hydrodynamic motions) and the lack of local thermodynamic equilibrium in stellar atmospheres. Such estimates are even more difficult for the spectra of galaxies, which comprise many stars that have different masses, ages and chemical compositions. In galaxies, spectral absorption lines are broadened because of the motion of stars and weakened by emissions of young stars or AGN. A further problem for spectroscopy is that in galaxies younger than 1 billion years, about 40% of the light comes from asymptotic giant branch (AGB) stars⁸, which are not well understood.

Despite these difficulties, Kriek and colleagues use spectroscopy to find a high α/Fe ratio in a galaxy that has a relatively high metallicity — the ratio of Fe to hydrogen is half

that of the Sun's. The combination of these two abundance ratios has never been seen in nearby stars: stars in the Galactic halo have high α/Fe ratios at less than one-tenth of the Sun's metallicity, those in the Galactic disk have lower α/Fe ratios at higher metallicities, and those in the Galactic bulge have high α/Fe ratios but only at one-third of the Sun's metallicity⁹.

What studies should be carried out to understand Kriek and collaborators' galaxy spectra? First, high-resolution spectra of metal-rich and α -enhanced stars — as in the authors' galaxy — need to be obtained for the Milky Way and should be compared with the authors' galaxy spectra. Such stars could be discovered using ongoing galactic-archaeology surveys such as the Apache Point Observatory Galactic Evolution Experiment (APOGEE) in New Mexico. Second, it is important to understand the evolution and spectra of AGB stars, both observationally and theoretically. If these stars are members of binary systems, their evolution and spectra would be very different. Finally, because the time at which type Ia supernovae start to occur is key to interpreting the authors' observations, it is crucial to study the progenitor systems, which are also binary systems, that cause these supernovae. This is closely related to the evolution of the Universe itself — type Ia supernovae were used in the 2011 Nobel-prizewinning discovery that the Universe's expansion is accelerating^{10,11}.

If the α/Fe ratio in the authors' high-redshift

galaxy is confirmed, this would challenge our current models of galaxy formation and raise many questions. For example, what kind of physical process can stop star formation on such a short timescale? Star formation in this galaxy might have been boosted and then quenched so that α/Fe is higher than usual. Alternatively, material produced in type Ia supernovae might have been efficiently removed by galactic winds. If this ejection was caused by AGN, what is the origin of the corresponding supermassive black holes?

Finally, how did the authors' galaxy evolve after star formation stopped? Is there a present-day counterpart of this galaxy, and how common are such galaxies? Even with the largest telescopes and the best detectors that astronomers have at present, it is difficult to measure the elemental abundances of a large sample of galaxies at high redshifts. However, such measurements will be possible in the near future with the James Webb Space Telescope and with extremely large (25–40 metres in diameter) ground-based telescopes. Meanwhile, theorists will need to predict elemental abundances across cosmic time in the quest to understand the formation and evolution of galaxies. ■

Chiaki Kobayashi is in the Centre for Astrophysics Research, School of Physics, Astronomy and Mathematics, University of Hertfordshire, Hatfield AL10 9AB, UK. e-mail: c.kobayashi@herts.ac.uk

1. Kriek, M. *et al.* *Nature* **540**, 248–251 (2016).
2. White, S. D. M. & Frenk, C. S. *Astrophys. J.* **379**, 52–79 (1991).
3. Hoyle, F. *Astrophys. J. (Suppl.)* **1**, 121–146 (1954).
4. Nomoto, K., Kobayashi, C. & Tominaga, N. *Annu. Rev. Astron. Astrophys.* **51**, 457–509 (2013).
5. Ji, A. P., Frebel, A., Chiti, A. & Simon, J. D. *Nature* **531**, 610–613 (2016).
6. Taylor, P. & Kobayashi, C. *Mon. Not. R. Astron. Soc.* **448**, 1835–1846 (2015).
7. Heckman, T. M. & Best, P. N. *Annu. Rev. Astron. Astrophys.* **52**, 589–660 (2014).
8. Maraston, C. *Mon. Not. R. Astron. Soc.* **362**, 799825 (2005).
9. Johnson, C. I., Rich, R. M., Kobayashi, C., Kunder, A., & Koch, A. *Astron. J.* **148**, 67 (2014).
10. Perlmutter, S. *et al.* *Nature* **391**, 51–54 (1998).
11. Riess, A. G. *et al.* *Astron. J.* **116**, 1009–1038 (1998).

NEURODEGENERATIVE DISORDERS

Neural synchronization in Alzheimer's disease

Electrical oscillations generated by neural circuits are disrupted in Alzheimer's disease. Restoring these oscillations in mouse models activates immune cells to clear disease-associated amyloid- β protein from the brain. [SEE ARTICLE P.230](#)

LIVIU ARON & BRUCE A. YANKNER

An intriguing feature of the brain is its generation of electrical oscillations through the synchronized activity of networks of neurons. The frequencies of such brain rhythms can span several orders of magnitude, from slow oscillations in the delta (0.5–3 hertz) range to gamma (30–90 Hz) and ultrafast (90–200 Hz) ranges. These rhythms have been implicated in fundamental neural processes, such as attention, sensory perception, learning and memory¹. Moreover, disruption of gamma oscillations has been observed in several neurological conditions, including brain trauma, schizophrenia and Alzheimer's disease². On page 230, Iaccarino *et al.*³ present evidence in mice that disruption of gamma oscillations might contribute to the accumulation of amyloid- β protein in the brain — a hallmark of Alzheimer's disease.

The brains of people with Alzheimer's show desynchronized electrical activity and loss of oscillatory activity, particularly of gamma-frequency oscillations^{2,4,5}. Mice carrying genetic mutations that cause Alzheimer's in humans also show altered electrical activity⁶ and, in one model, restoration of normal gamma oscillations has been shown to reduce memory deficits⁷. It is not known, however, whether these changes in neural activity have a causal role in the biological changes underlying disease progression or represent secondary phenomena.

Iaccarino and colleagues explored the effects of gamma activity on amyloid- β (A β) accumulation in a mouse model of Alzheimer's disease known as 5XFAD. The authors showed that these mice exhibit low gamma activity, similar to other mouse models of the disease^{7,8}. To restore this activity, mice were genetically engineered to express the light-activated ion-channel protein channelrhodopsin in neurons called fast-spiking parvalbumin interneurons, which are involved in the generation of

gamma oscillations⁹. Activation of channelrhodopsin using a fibre-optic light source implanted in the hippocampus region of the animals' brains induced synchronous neuron firing and gamma oscillations. Remarkably, this led to a major reduction in the accumulation of disease-associated A β deposits in the hippocampus.

How could augmentation of gamma oscillations have such a major effect? The first clue came when Iaccarino and co-workers analysed RNA transcripts in the hippocampus of 5XFAD mice, and found that gamma stimulation increased the expression of many genes involved in the function of resident immune cells called microglia. Some of the most affected genes were associated with phagocytosis — the process by which immune cells take up extracellular material for clearance and degradation. Furthermore, gamma stimulation caused microglia to adopt an 'activated' shape

and induced increased intracellular A β levels, consistent with A β phagocytosis (Fig. 1). Biochemical experiments suggested that gamma stimulation also altered the processing of amyloid precursor protein, indicating reduced generation of A β . Thus, gamma stimulation might have a bimodal effect, both reducing A β generation and increasing its clearance by microglia.

The authors then explored a non-invasive approach to stimulating natural production of gamma oscillations. Previous work¹⁰ showed that visual stimulation using flickering light at specific frequencies induced gamma oscillations in the brain's visual cortex. In an elegant series of experiments, the authors found that exposing 5XFAD mice to a 40-Hz flickering light for just 1 hour augmented gamma oscillations and markedly reduced A β levels for 12 to 24 hours. Repeating the treatment for 7 days reduced the load of harmful A β deposits called plaques in the visual cortex by about 60%, suggesting that there could be a longer-term effect. No effect on A β levels was seen after exposure to other frequencies or a random-frequency light flicker. A β deposition is therefore highly sensitive to specific patterns of neural activation.

Light-flicker treatment induced gamma oscillations and reduced A β levels in ageing wild-type mice, consistent with a physiological role for gamma stimulation in regulating A β metabolism. Interestingly, the treatment also reduced the build-up of tau proteins in a mouse model of frontotemporal dementia, suggesting that gamma activity might have

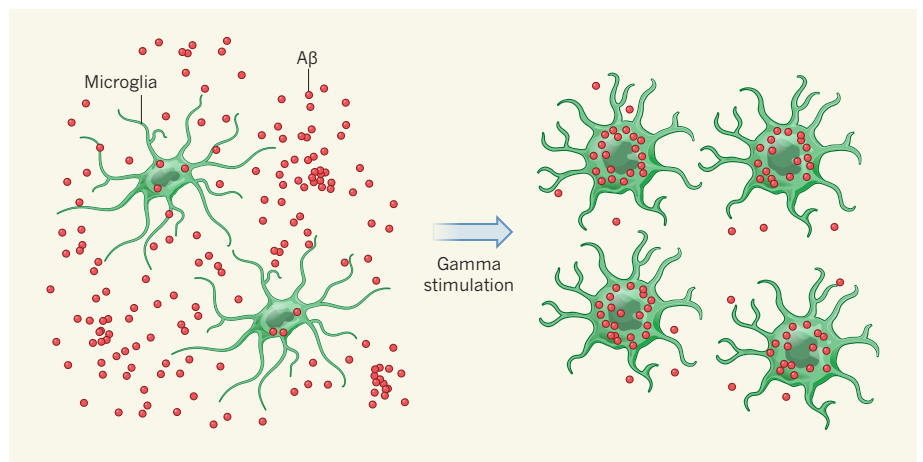


Figure 1 | Gamma oscillations stimulate the clearance of amyloid- β protein deposits. Abnormal aggregation of amyloid- β (A β) protein in the brain is associated with Alzheimer's disease. A β aggregates might accumulate and promote neurodegeneration in part because immune cells called microglia cannot effectively clear the protein. In addition, synchronized patterns of electrical activity in the brain known as gamma oscillations are disrupted in Alzheimer's. Iaccarino *et al.*³ restored gamma oscillations in a mouse model of the disease. Such gamma stimulation led to recruitment of microglia to sites of A β deposition. The microglia adopted an activated shape, and consequently engulfed and degraded A β .

broad effects on protein homeostasis in the brain.

In the past ten years, dysfunction in neural networks has emerged as a potential contributor to the biology and clinical symptoms of Alzheimer's disease¹¹. First, it was observed that neural activity could regulate local concentrations of A β in the interstitial fluid that surrounds cells in the brain¹². Subsequently, the topography of neural connections was shown to direct the spread of tau and A β protein aggregates in mouse models of neurodegenerative disease^{13,14}. Moreover, circuit connectivity between the hippocampus and the cortex has been proposed to be a primary driver of the stepwise spread of intracellular tau aggregates called neurofibrillary tangles between these brain regions in people with Alzheimer's¹⁵. Iaccarino and colleagues suggest that not only connectivity, but also patterns of neural activity — in particular, gamma oscillations — play a part. In this model, dysfunctional neural-network activity promotes the accumulation and spread of misfolded proteins that, in turn, cause further circuit disruption in a harmful positive feedback loop.

The observation that microglia clear A β in response to specific patterns of neural activity raises several questions. Does synchronous neural firing result in the secretion of factors that attract microglia and activate their phagocytic activity? Could such factors be used as therapeutic agents? Does this mode of microglial stimulation have a role in the function and maintenance of the synaptic connections between neurons in healthy brains? Recent reports^{16,17} suggest that microglia mediate synaptic pruning, which occurs as a normal part of brain development and might become dysfunctional in Alzheimer's disease. The process might involve phagocytosis and turnover of synaptic structural components — a function that could be co-opted for clearance of A β and other protein aggregates.

Treating neurodegenerative disorders by non-invasive modulation of neural networks is an intriguing prospect. The present study used light to stimulate gamma oscillations in the visual cortex, which is relatively unaffected by Alzheimer's disease. It will be important to determine whether other approaches can stimulate gamma oscillations more globally, in brain regions that are affected in Alzheimer's and other disorders. For example, behavioural interventions such as meditation have been shown to increase gamma oscillations¹⁸. Electrical stimulation of deep brain regions — an effective approach in drug-resistant Parkinson's disease — could potentially be adapted to stimulate gamma oscillations in specific brain regions. These are just two of many potential therapeutic approaches likely to arise from a greater understanding of the role of neural networks in neurodegenerative disorders. ■

Liviu Aron and Bruce A. Yankner are in the

Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.
e-mail: bruce_yankner@hms.harvard.edu

1. Wang, X.-J. *Physiol. Rev.* **90**, 1195–1268 (2010).
2. Uhlhaas, P. J. & Singer, W. *Neuron* **52**, 155–168 (2006).
3. Iaccarino, H. F. *et al. Nature* **540**, 230–235 (2016).
4. Koenig, T. *et al. Neurobiol. Aging* **26**, 165–171 (2005).
5. Pijnenburg, Y. A. L. *et al. Clin. Neurophysiol.* **115**, 1332–1339 (2004).
6. Palop, J. J. *et al. Neuron* **55**, 697–711 (2007).
7. Verret, L. *et al. Cell* **149**, 708–721 (2012).
8. Gillespie, A. K. *et al. Neuron* **90**, 740–751 (2016).
9. Sohail, V. S., Zhang, F., Yizhar, O. & Deisseroth, K. *Nature* **459**, 698–702 (2009).

10. Fries, P., Nikolić, D. & Singer, W. *Trends Neurosci.* **30**, 309–316 (2007).
11. Palop, J. J. & Mucke, L. *Nature Neurosci.* **13**, 812–818 (2010).
12. Cirrito, J. R. *et al. Neuron* **48**, 913–922 (2005).
13. Brettschneider, J., Del Tredici, K., Lee, V. M.-Y. & Trojanowski, J. Q. *Nature Rev. Neurosci.* **16**, 109–120 (2015).
14. Pooler, A. M. *et al. Alzheimers Res. Ther.* **5**, 49 (2013).
15. Yankner, B. A., Lu, T. & Loerch, P. *Annu. Rev. Pathol.* **3**, 41–66 (2008).
16. Hong, S., Dissing-Olesen, L. & Stevens, B. *Curr. Opin. Neurobiol.* **36**, 128–134 (2015).
17. Hong, S. *et al. Science* **352**, 712–716 (2016).
18. Lutz, A., Greischar, L. L., Rawlings, N. B., Ricard, M. & Davidson, R. J. *Proc. Natl Acad. Sci. USA* **101**, 16369–16373 (2004).

In Retrospect

Forty years of linking orbits to ice ages

In 1976, it was demonstrated that tiny wobbles in Earth's orbit led to the great ice-age cycles of the past few million years. This finding had wide implications for climate science and the details remain hotly debated today.

MARK MASLIN

Forty years ago, Hays, Imbrie and Shackleton¹ published one of the most influential papers in the study of Earth's past climate. The authors showed that variations in Earth's orbital path around the Sun were the pacemaker of the great ice ages in the Quaternary period (the current geological period, which began about 2.6 million years ago). Their study is a beautiful example of how three exceptional scientists worked together to robustly test a controversial theory, thereby allowing us to understand the fundamental causes of the ice ages and to peer into the future to see whether humanity has disrupted these natural cycles.

Milutin Milanković was a brilliant Serbian mathematician and climatologist who in 1941 postulated that wobbles in Earth's orbit changed the distribution of solar energy on the planet's surface, pushing Earth in or out of an ice age² — a process called orbital forcing. He suggested that insolation (the amount of incoming solar radiation) at a latitude of 65°N, just south of the Arctic Circle, was critical. At this latitude, insolation can vary by 25% (from 430 to 560 watts per square metre)². Milanković argued that, when insolation was reduced during the summer, some of the ice in this region could survive. Each year, the ice would build up to eventually produce an ice sheet.

In 1976, three gifted scientists joined forces to test Milanković's theory, using long-term climate records obtained by analysing marine

sediments. James Hays founded and led the international CLIMAP project, which used fossil assemblages to estimate past sea surface temperatures and brought the dream team together. Nick Shackleton was a master of stratigraphy and provided oxygen-isotope records, which showed past global ice volumes. John Imbrie was the expert in time-series analysis, particularly spectral analysis, which showed that the climate records matched calculations of high-latitude insolation. Remarkably, the authors discovered that these records

“Hays, Imbrie and Shackleton legitimized what was to become one of the most powerful tools in stratigraphy.”

contained the same temporal cycles — known as Milankovitch cycles — as three parameters that describe Earth's orbit: eccentricity, obliquity and precession (Fig. 1).

Eccentricity describes the shape of Earth's orbit around the Sun, and varies from nearly circular to elliptical, in part because of Jupiter's gravity. Obliquity is the tilt of Earth's axis of rotation with respect to the plane of its orbit, and directly affects the intensity of the seasons. Finally, precession is the most complicated type of variation because it alters the distance between Earth and the Sun during each season, and has two components — Earth's rotational axis precesses (rotates) owing to tidal forces exerted by the Sun and the Moon on the solid Earth, and Earth's orbital path itself precesses around the Sun.

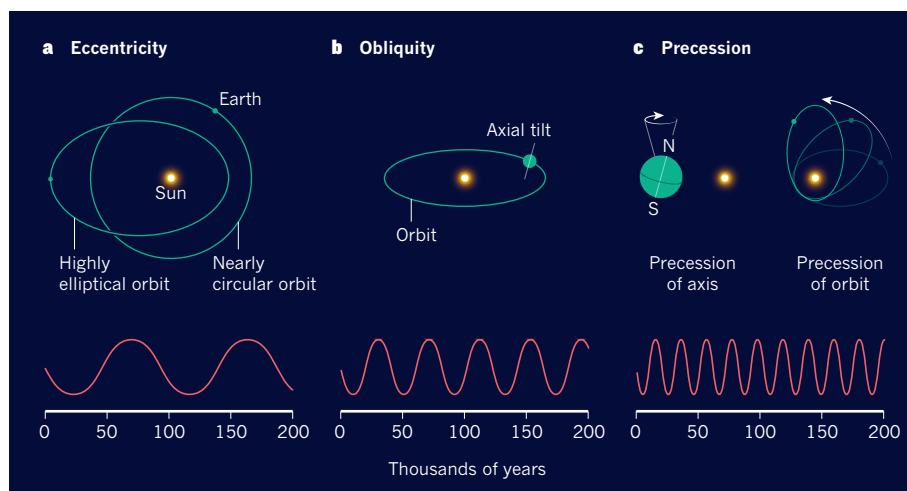


Figure 1 | Milankovitch cycles. In 1976, Hays, Imbrie and Shackleton¹ provided the first clear evidence that variations in Earth's orbit around the Sun dramatically change our planet's climate, a concept known as the Milankovitch theory². There are three types of variation: eccentricity, obliquity and precession^{1,2}. **a**, Eccentricity describes the shape of Earth's orbit around the Sun, and varies from being almost a circle to an ellipse, with a period of about 96,000 years⁶. **b**, Obliquity is the tilt of Earth's axis of rotation with respect to the plane of its orbit, and it oscillates with a period of about 41,000 years^{6,8}. **c**, Both Earth's rotational axis and its orbital path precess (rotate) over time — the combined effects of these two components and the eccentricity produce an approximately 21,000-year cycle^{6,7}.

By demonstrating the clear link between orbital forcing and Earth's past climate, Hays, Imbrie and Shackleton legitimized what was to become one of the most powerful tools in stratigraphy. For example, reliable age models have been constructed for climate records covering at least the past 5 million years by tuning the orbital parameters to the ice-age cycles³. Such age models can be applied to any long-term palaeoclimate record, allowing marine and land records to be compared.

In addition, the various effects of the three orbital parameters have been used to study orbital forcing at different latitudes. Obliquity has a strong influence at high latitudes, whereas precession has a significant impact on seasonality in the tropics — precession has been linked to the rise and fall of the African rift-valley lakes, and even to our own evolution⁴. Evidence for the orbital forcing of climate has now been found as far back as 1.4 billion years ago, in the Proterozoic eon⁵.

Hays, Imbrie and Shackleton clearly set out the limitations of their study and presented the scientific community with a range of challenges, many of which remain today. In particular, the authors recognized that variations in the orbital parameters did not cause the ice-age cycles, but rather paced them. Any given combination of parameters can be associated with many different climates — for example, Earth's orbital configuration today is similar to that of 18,000 years ago, when a 3-kilometre-thick ice sheet covered North America. Feedback mechanisms take the small changes in insolation that are driven by the orbital parameters and push Earth into or out of an ice age. Therefore, the next step was to understand the

relative importance of the feedbacks involving the ice sheets, oceans and atmosphere; this led to the discovery that greenhouse gases had a pivotal role in controlling past climate.

The authors' work also recognized what is known as the 100,000-year problem. Before 1 million years ago, ice ages occurred roughly every 41,000 years owing to variations in Earth's obliquity⁶. This makes climatological sense, because Earth's axial tilt directly controls how warm or cold the summers are in the Northern Hemisphere. But the past 8 ice-age cycles had a longer period of 100,000 years⁷, which is similar to the period associated with eccentricity. In terms of forcing, eccentricity is by far the weakest of the three orbital parameters. Therefore, if eccentricity were responsible for such 100,000-year cycles, there would need to be a complicated 'nonlinear' amplification effect by Earth's climate system.

However, the similarity between the two periods turned out to be an artefact of spectral analysis⁶ — although the previous 8 ice-age cycles lasted for about 100,000 years on average, they ranged in length from 80,000 to 120,000 years. With the realization that eccentricity is not the major driving force, a debate has emerged as to whether precession or obliquity controlled the timing of the most recent ice-age cycles. Some researchers argue that the deglaciations occurred every four or five precessional cycles^{6,7}, others that it was every second or third obliquity cycle⁸, and some argue that it was a combination of the two⁹. The debate started 40 years ago and still rages today.

The authors' work also provided a tool with which to investigate the future of Milankovitch cycles. It has been suggested



50 Years Ago

At the neuromuscular junction¹, as well as at neuronal synapses², calcium plays an essential and direct part in the process whereby depolarization of the presynaptic nerve terminal leads to release of the transmitter substance. It is also known that a nerve impulse releases the transmitter in quantal form, that is, in discrete multimolecular amounts of a fairly standard size ... The question is raised whether other ions can replace calcium in this process and if so is the transmitter still released in quantal form? ... Fig. 1 shows sample records from one experiment where calcium and strontium were applied simultaneously to different synaptic spots on the same end-plate ... These experiments show that strontium can replace calcium in the process of transmitter release by a nerve impulse and that the transmitter is still released in quantal fashion.

From *Nature* 10 December 1966

100 Years Ago

Prof. J. W. Gregory, in his presidential address to the Geological Society of Glasgow, gave an account of the chief sources of the world's supply of phosphates, in the course of which he pointed out that an instructive lesson in the conservation of mineral resources was to be learnt from this subject. He showed that Britain has but limited supplies of natural phosphates, and these were being left unworked owing to the introduction of cheaper and richer products from foreign deposits. Prof. Gregory has done valuable service in again directing attention to our supply of phosphates, and it is clear that, from the point of view both of the natural and the artificial phosphate supply, the question is one of vital importance to our great agricultural interests.

From *Nature* 14 December 1916

that small increases in greenhouse gases due to the expansion of agriculture that started 8,000 years ago¹⁰ have, in fact, delayed the next ice age¹¹. Moreover, if greenhouse-gas emissions continue to grow, the next ice age might be postponed for at least half a million years¹². Understanding orbital forcing is therefore relevant to contemporary debates about the Anthropocene — a proposed geological unit that is defined by human activity. If we have merely delayed the next ice age, we will still be in the Quaternary period, and the Anthropocene can be defined as an epoch, a subdivision of time below a period. But if we have stopped the ice ages, we will have entered

the Anthropocene period¹³, marking a much larger change in Earth's climate system. Therefore, unravelling the causes of the great ice ages is pivotal to understanding both the past and our future. ■

Mark Maslin is in the Department of Geography, University College London, London WC1E 6BT, UK.
e-mail: m.maslin@ucl.ac.uk

1. Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194**, 1121–1132 (1976).
2. Milanković, M. M. *Canon of Insolation and the Ice-Age Problem* (R. Serb. Acad., 1941).
3. Lisiecki, L. E. & Raymo, M. E. *Paleoceanography*

20, PA1003 (2005).

4. Maslin, M. A. *et al. Quat. Sci. Rev.* **101**, 1–17 (2014).
5. Zhang, S. *et al. Proc. Natl Acad. Sci. USA* **112**, E1406–E1413 (2015).
6. Maslin, M. A. & Brierley, C. M. *Quat. Int.* **389**, 47–55 (2015).
7. Ridgwell, A., Watson, A. J. & Raymo, M. E. *Paleoceanography* **14**, 437–440 (1999).
8. Huybers, P. J. *Nature* **480**, 229–232 (2011).
9. Huybers, P. & Wunsch, C. *Nature* **434**, 491–494 (2005).
10. Ruddiman, W. F. *Clim. Change* **61**, 261–293 (2003).
11. Tzedakis, P. C., Channell, J. E. T., Hodell, D. A., Kleiven, H. F. & Skinner, L. C. *Nature Geosci.* **5**, 138–141 (2012).
12. Archer, D. & Ganopolski, A. *Geochem. Geophys. Geosyst.* **6**, Q05003 (2005).
13. Lewis, S. L. & Maslin, M. A. *Nature* **519**, 171–180 (2015).

BIOMEDICINE

Replacing the cell's power plants

Nuclear DNA from human eggs that harbour mutations in the DNA of organelles called mitochondria has been successfully transferred to donor eggs, bringing the prospect of therapy for mitochondrial diseases a step closer. SEE LETTER P.270

ERIC A. SHOUBRIDGE

A revolution in DNA-sequencing technology in the past few years has allowed relatively easy identification of the underlying causes of many genetic diseases. However, preventing transmission of disease-causing defects remains a formidable challenge. Our cells contain two genomes — one in the nucleus and another much smaller, semi-autonomous one in organelles called mitochondria, which are essential for the production of energy. Diseases caused by mutations in mitochondrial DNA (mtDNA) affect about 1 in 5,000 people¹, have an extraordinarily broad spectrum of symptoms and are currently untreatable. Kang *et al.*² show on page 270 that it is possible to prevent harmful mutations in human mtDNA from being transmitted to offspring, using a mitochondrial replacement technique.

We inherit all of our mitochondria, and so all of our mtDNA, from our mothers. Female eggs carry a few hundred thousand copies of mtDNA, and most cells in an adult contain hundreds to thousands of these copies. In individuals with mtDNA disease there is usually a mixture of normal and mutated mtDNAs, and the severity of disease generally correlates with the proportion of mtDNA copies that carry the disease-causing mutation.

This proportion can vary widely between eggs from the same woman. One way to avoid the transmission of mutations is to combine *in vitro* fertilization (IVF) with

genetic diagnosis to identify and implant only embryos harbouring normal mtDNAs. Although this approach has been successful³, it is not always possible to recover suitable embryos.

Another strategy might be to swap the

mitochondria containing mutated mtDNA for healthy ones (Fig. 1). This is the idea behind mitochondrial replacement techniques (MRTs). In one MRT, known as meiotic-spindle transfer, nuclear DNA from the egg of a 'carrier' mother harbouring mutated mtDNA would be transferred to a donor egg that lacks a nucleus and contains only normal copies of mtDNA. This egg would then be fertilized *in vitro* and implanted. A variation on the theme is pronuclear transfer, in which the nuclei from the sperm and egg would be transferred to a nucleus-free donor egg immediately after fertilization, before the two have fused. In both cases, the reconstituted embryo would carry the same genes as its biological parents, except for the few on mtDNA that come from an unrelated donor.

The group that conducted the current study has previously performed proof-of-principle

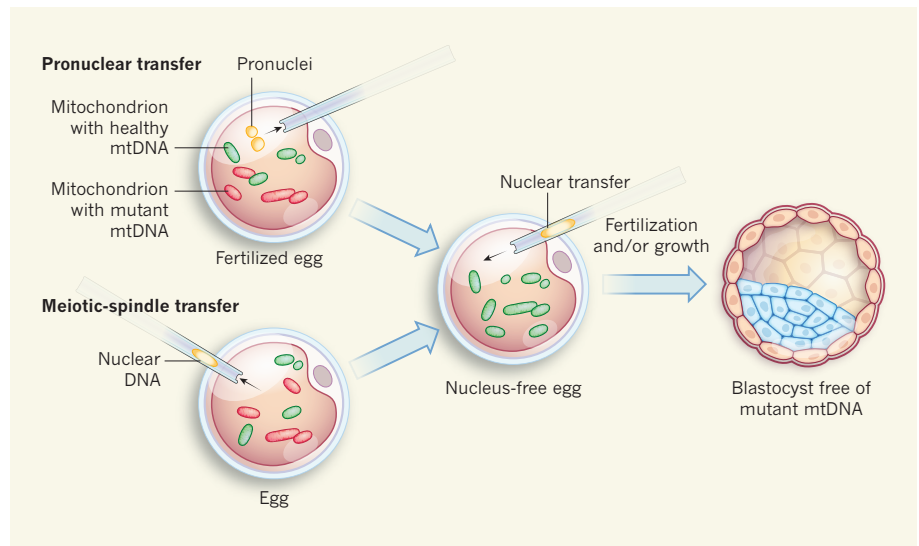


Figure 1 | Putative techniques for mitochondrial replacement. Mutations in the DNA of organelles called mitochondria cause disease. A possible therapy would involve replacing mutated mitochondrial DNAs (mtDNAs) with healthy ones. One such therapy would be transfer of pronuclei — sperm and egg nuclei post-fertilization, before the two fuse — from an egg that harbours some mutated mtDNAs into a donor cell that has healthy mtDNA, from which the nucleus has been removed. In another technique, known as meiotic-spindle transfer, nuclear DNA from an egg harbouring some mutant mtDNAs would be transferred into a nucleus-free egg, which would subsequently be fertilized. Kang *et al.*² provide evidence that meiotic-spindle transfer can produce healthy embryos at the blastocyst stage of development that are free of mutant mtDNA.

tests to show that meiotic-spindle transfer can produce healthy macaque offspring⁴ and human embryos that develop normally up to the blastocyst stage at five days of development⁵. Earlier this year, another group had a similar success with pronuclear transfer⁶. However, these studies used carrier eggs that contained normal mtDNA only.

In the current study, Kang *et al.* turned to eggs from women carrying single-nucleotide mtDNA mutations. The authors successfully used meiotic-spindle transfer to produce embryos that developed to the blastocyst stage and that carried a virtually undetectable proportion of mutant mtDNAs. So, it seems that the precise timing of nuclear transfer (before or after fertilization) is unimportant. However, because pronuclear transfer has not been attempted using eggs from women carrying mutant mtDNA, a direct comparison would be required to confirm this speculation.

Concerns have been raised about possible nuclear-mitochondrial incompatibilities resulting from MRT — the idea that certain combinations of nuclear and mitochondrial genomes could have adverse effects on cell fitness. However, evidence to support such concerns is based on studies of inbred mice^{7,8}, which have no genetic variation in their nuclear genomes. Human populations, by contrast, are highly heterogeneous, and so paternal genes find themselves in a potentially novel mtDNA environment with each generation. Thus, it seems unlikely that any kind of tight coupling would have evolved between nuclear and mitochondrial DNA in humans. Indeed, Kang and colleagues show that embryo development is independent of the genetic distance between the mtDNA of the carrier and that of the donor, corroborating observations⁵ in a macaque MRT experiment that used donors and carriers with highly divergent mtDNAs.

Another concern is the fate of the 2% or so of mtDNA copies that hitch-hike with the nucleus from the carrier egg into the reconstructed embryo. To investigate the potential consequences of such hitch-hiking, Kang *et al.* analysed embryonic stem cells that they isolated from the reconstructed embryos and grew in culture. Their analyses showed that, in rare instances, donor mtDNAs can be outcompeted by even small proportions of carrier mtDNA after a large number of cell divisions — in line with the results of two previous studies^{6,9}.

Two possible mechanisms could be in play here: genetic drift (random changes to genetic make-up that occur over generations) and replicative advantage. If the changes seen in mtDNA proportions are due entirely to genetic drift, the probability that a rare mutation in a mixed mtDNA population will become predominant is directly proportional to its initial frequency — as has been demonstrated *in vivo* for cells in crypt structures in the colon, which are continuously regenerated from a stem-cell

population¹⁰. This probably would not be much of a worry for therapeutics, because most cells would lose the rare mtDNAs carrying the mutation.

But Kang and colleagues' study shows that, in some cases, the carrier mtDNA could have a replicative advantage over mtDNA from the donor, probably owing to sequence variation in the region of mtDNA that regulates the replication rate of the mitochondrial genome. This phenomenon could perpetuate a systematic return to a majority of carrier mtDNA. As the authors suggest, it should be possible to select donors on the basis of relative efficiencies of mtDNA replication when studying the process in model systems. But the problem cannot be directly studied in human embryos, because mtDNA replication is stalled until after the blastocyst stage. It would, however, be useful to investigate the phenomenon in other mammals, where it is possible to study replication of mtDNA throughout early embryonic development.

Aside from concerns about the fate of hitch-hiking mtDNA from the carrier egg, what is holding back widespread application of MRTs? The procedures are technically challenging and so not likely to be licensed widely in IVF clinics. It is also currently unclear how many carrier or donor eggs would have to be used

to reliably produce healthy reconstructed embryos for implantation. A recent report of a baby boy born after an MRT (see *Nature* <http://doi.org/btr9>; 2016), performed by a US physician in Mexico, made headlines, highlighting the need for carefully supervised clinical trials. Legislation has so far legalized the procedure in a single jurisdiction — the United Kingdom. Until that changes, it seems unlikely that many women will benefit. ■

Eric A. Shoubridge is in the Department of Human Genetics, Montreal Neurological Institute of McGill University, Montreal, Quebec H3A 2B4, Canada.
e-mail: eric@ericpc.mni.mcgill.ca

- Schaefer, A. M. *et al.* *Ann. Neurol.* **63**, 35–39 (2008).
- Kang, E. *et al.* *Nature* **540**, 270–275 (2016).
- Steffann, J. *et al.* *J. Med. Genet.* **43**, 244–247 (2006).
- Tachibana, M. *et al.* *Nature* **461**, 367–372 (2009).
- Tachibana, M. *et al.* *Nature* **493**, 627–631 (2013).
- Hyslop, L. A. *et al.* *Nature* **534**, 383–386 (2016).
- Latorre-Pellicer, A. *et al.* *Nature* **535**, 561–565 (2016).
- Moreno-Loshuertos, R. *et al.* *Nature Genet.* **38**, 1261–1268 (2006).
- Yamada, M. *et al.* *Cell Stem Cell* **18**, 749–754 (2016).
- Jenuth, J. P., Peterson, A. C. & Shoubridge, E. A. *Nature Genet.* **16**, 93–95 (1997).

This article was published online on 30 November 2016.

STEM CELLS

Aspiring to naivety

Human stem cells that can give rise to every cell type in the body are major players in biomedical research. A molecular analysis of human embryos might help to make these cultured cells more authentic imitations of their *in vivo* counterparts.

IDO SAGI & NISSIM BENVENISTY

Human embryonic stem cells have revolutionized biomedical research, thanks to the fact that they are pluripotent: able to differentiate into every cell lineage of the body. These cells are grown in culture, after being isolated at the blastocyst stage of early development from an embryo produced by *in vitro* fertilization. But under conventional culture conditions, embryonic stem cells are not considered to be naive pluripotent stem cells (PSCs) — those identical to cells of the blastocyst from which they are derived. Instead, they are regarded as primed PSCs, resembling more-mature cells that are poised to differentiate¹. Writing in *Cell Stem Cell*, Theunissen *et al.*² describe a set of criteria that define naive human PSCs, and so take an important step towards generating truly naive cells in culture. Such an ability could provide opportunities to study early developmental processes, which in turn could make it possible

to optimize cell-differentiation protocols for disease modelling and therapy.

In the past few years, several groups of researchers have defined culture conditions that seem to sustain human PSCs in a blastocyst-like naive state, either by conversion from the primed state or by direct isolation from embryos^{3–6} (Fig. 1). However, it has remained unclear exactly how similar these 'naive' cells are to *in vivo* blastocyst cells. The different cocktails of growth factors and small molecules used by each research group produced human cells that had some properties of naive mouse pluripotency, and that varied in their resemblance to the human blastocyst in terms of gene expression⁷. However, embryonic development is substantially different in humans and mice with regard to timing, morphology and at the molecular level. As such, comparisons with mouse pluripotency might be a misleading way to gauge the state of human cells. To overcome this obstacle, Theunissen *et al.* turned to human embryos.

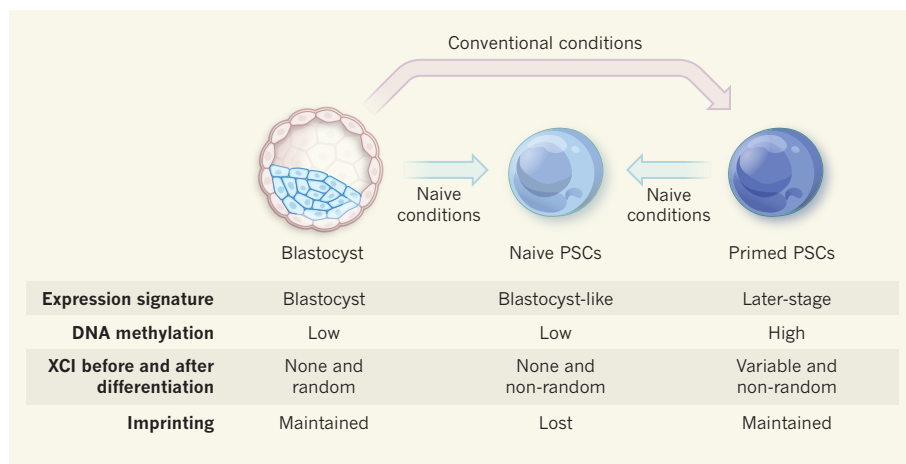


Figure 1 | Molecular characterization of naive human stem cells. Human pluripotent stem cells (PSCs) are isolated from an embryo at the blastocyst stage of development. These cells, grown under conventional culture conditions, are primed to differentiate into every lineage in the body. It is desirable to identify culture conditions that sustain human PSCs in a less-mature, blastocyst-like naive state, following either isolation from the embryo or their conversion from primed cultures. Theunissen *et al.*² compared molecular properties of human blastocysts and PSCs to define criteria for naive human pluripotency. Naive PSCs resemble blastocyst cells but not primed cells in several ways: expression of genes and transposable DNA elements; levels of a molecular modification to DNA called methylation; and the lack of X-chromosome inactivation (XCI) in female cells. But, unlike blastocyst cells, which undergo random XCI following differentiation, the process is not random in naive cells. Furthermore, imprinting (the inactivation of one copy of a gene inherited from a given parent) is lost in naive cells.

The authors began by devising an approach to robustly distinguish between different embryonic stages on the basis of RNA expression. Roughly half of human DNA derives from mobile sequences called transposable elements, which have become embedded in the genome over the course of evolution. These abundant sequences often produce RNA in a cell-type-specific manner, and Theunissen and colleagues exploited this characteristic to define high-resolution and stage-specific signatures of transposable-element expression. They found that the genes expressed in naive cells (generated using up-to-date protocols) mostly overlapped with genes expressed in the blastocyst, but the cells' transposable-element profile corresponded, unexpectedly, to both the blastocyst and an earlier developmental stage. The analysis further pointed to human-specific families of related transposable elements whose expression is characteristic of either naive or primed cells, providing a possible explanation for some differences between pluripotent states in mice and humans.

Theunissen *et al.* next focused on a widespread chemical modification to DNA called methylation, which is associated with the regulation of gene expression. In early embryos, genome-wide DNA methylation levels decrease dramatically compared with those in the egg and sperm, reaching a minimum at the blastocyst stage⁸. Yet methylation is maintained at imprinted sequences — regions in which the copy of a gene inherited from one parent is expressed, and the copy from the other parent is silenced. Primed PSCs typically have heavily methylated genomes and maintain their

imprints^{8,9}. Reassuringly, the authors found that naive cells showed low average methylation levels, comparable with those in the blastocyst. As expected, methylation levels increased when the cells were induced to differentiate.

However, the naive cells irreversibly lost methylation at most imprinted regions, leading to aberrant gene expression. These results mirror those of a recent study¹⁰, which concluded that patterns of methylation, unlike average levels, are poorly correlated between naive and blastocyst cells. This implies that

“Understanding human pluripotency is essential for fully exploiting its potential.”

mechanisms other than those that operate during development influence DNA methylation in naive cells *in vitro*. Because loss of imprinting is associated with various human diseases (such as Prader–Willi and Beckwith–Wiedemann syndromes), resolving this issue is imperative if current protocols for generating naive PSCs are to be used for therapy.

A major advance made by Theunissen and colleagues relates to X-chromosome inactivation (XCI) — the random silencing of one of the two X chromosomes in female cells, which ensures that gene-expression levels from the X chromosome are the same as those in males. XCI is initiated during differentiation by the non-coding RNA *XIST*. But in the human blastocyst, both X chromosomes are active despite *XIST* expression, and another, unknown mechanism acts to balance

expression levels¹¹. Primed PSCs exhibit variable XCI states¹, but the authors found that both X chromosomes were activated despite *XIST* expression in naive cells, as in the blastocyst.

To explore XCI further, Theunissen *et al.* inserted genes encoding different fluorescent ‘marker’ proteins into the two X chromosomes. In the primed state, the fluorescent gene inserted into the inactive X chromosome remained silent, whereas the other gene was expressed. When the primed cells were converted to the naive state, the silent reporter became activated. However, when the researchers differentiated these naive cells, the chromosome that had been inactive when the cells were primed was always the one to be inactivated again, suggesting that the process does not mimic normal, random XCI, but reflects residual molecular memory.

Finally, the authors tested whether naive PSCs were functional, by trying to integrate them into an early-stage host embryo and analysing whether they became part of the developing embryo — a gold-standard test for functional mouse PSCs. Using human hosts is prohibited, so the authors instead integrated their human cells into more than 1,300 mouse embryos. However, human-cell integration was too inefficient to test the functionality of the naive human PSCs.

By defining clear criteria for naive pluripotency based on early human embryos, this study sets the groundwork for future experiments aiming to obtain human PSCs that match their *in vivo* counterparts. Understanding human pluripotency is essential for fully exploiting its potential. Further investigation into genome chemistry, nuclear organization and cellular metabolism will help to refine our definitions of pluripotent states and enhance our ability to mimic them *in vitro*. ■

Ido Sagi and Nissim Benvenisty are at The Azrieli Center for Stem Cells and Genetic Research, Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 91904, Israel.
e-mail: nissimb@mail.huji.ac.il

- De Los Angeles, A. *et al.* *Nature* **525**, 469–478 (2015).
- Theunissen, T. W. *et al.* *Cell Stem Cell* **19**, 502–515 (2016).
- Gafni, O. *et al.* *Nature* **504**, 282–286 (2013).
- Chan, Y.-S. *et al.* *Cell Stem Cell* **13**, 663–675 (2013).
- Theunissen, T. W. *et al.* *Cell Stem Cell* **15**, 471–487 (2014).
- Takahashi, Y. *et al.* *Cell* **158**, 1254–1269 (2014).
- Huang, K., Maruyama, T. & Fan, G. *Cell Stem Cell* **15**, 410–415 (2014).
- Smith, Z. D. *et al.* *Nature* **511**, 611–615 (2014).
- Rugg-Gunn, P. J., Ferguson-Smith, A. C. & Pedersen, R. A. *Nature Genet.* **37**, 585–587 (2005).
- Pastor, W. A. *et al.* *Cell Stem Cell* **18**, 323–329 (2016).
- Petropoulos, S. *et al.* *Cell* **165**, 1012–1026 (2016).

This article was published online on 30 November 2016.

Certified randomness in quantum physics

Antonio Acín^{1,2} & Lluís Masanes³

The concept of randomness plays an important part in many disciplines. On the one hand, the question of whether random processes exist is fundamental for our understanding of nature. On the other, randomness is a resource for cryptography, algorithms and simulations. Standard methods for generating randomness rely on assumptions about the devices that are often not valid in practice. However, quantum technologies enable new methods for generating certified randomness, based on the violation of Bell inequalities. These methods are referred to as device-independent because they do not rely on any modelling of the devices. Here we review efforts to design device-independent randomness generators and the associated challenges.

Because of its importance, much scientific effort is devoted to understanding when a given process generates ‘good’ randomness. A process is represented by a device, or black box, producing bits (Fig. 1), to which the user has access. What constitutes ‘good’ randomness may depend on the application, but here we are interested in the strongest definition: N bits are perfectly random if they are unpredictable, not only by the user of the device, but by any observer. This definition is satisfactory both from a fundamental and applied perspective. Although unpredictability by any observer may not be needed for some applications, such as Monte Carlo simulations, from a fundamental perspective it is difficult to argue that a process is random if there could exist an observer able to predict its outcomes. The requirement that the results should be unpredictable by any observer guarantees that the generated randomness is private: the user, by running the process in a secure location, can be certain that nobody knows the obtained results. Such private randomness can safely be used for cryptographic purposes.

According to our definition, the generation of randomness from scratch is impossible. This follows from the unfalsifiable hypothesis of the existence of a super-deterministic model in which everything, including the entire history of our Universe, was pre-determined in advance and known by the external observer. Therefore, any protocol for randomness generation must be based on some hypotheses or assumptions. The appropriateness of the assumptions is often debatable and strongly depends on the application. From a fundamental point of view, one random number generator (RNG) is better than another if it is based on fewer or weaker assumptions. However, adding more assumptions may not compromise the use of the random bits for a specific application, while it may simplify the experimental implementation and/or increase the efficiency.

Here we adopt a physics-based approach to randomness generation: the random numbers should be unpredictable by any physical observer, that is, any observer whose actions are constrained by the laws of physics. In particular, and according to the current understanding of nature, the device generating the random numbers and the device held by the potential adversary should obey the laws of quantum physics. Within this theory, the joint state of the user U and the adversary E (which includes the environment) describing N ideal random bits is

$$\rho_{UE} = \left(\frac{1}{2} |0\rangle\langle 0| + \frac{1}{2} |1\rangle\langle 1| \right)^{\otimes N} \otimes \rho_E \quad (1)$$

This state corresponds to N realizations of a perfect random bit, which takes the value 0 and 1 with equal probability and is totally uncorrelated with the state of the adversary, or environment, ρ_E . We say that

a given device generates ‘arbitrarily good randomness’ if its output is undistinguishable from the ideal state given in equation (1), up to some controllable small error.

A fundamental issue when considering randomness generation is certification: how can the user certify that the numbers produced by their device are random? According to equation (1), the random bits should follow a uniform probability distribution and be uncorrelated to the environment. Concerning the first point, the standard method for certifying randomness consists of running statistical tests¹ on sequences generated by the device. However, it is unclear what passing these tests means and, in fact, it is impossible to certify with finite computational power that a given sequence is random.

The certification of privacy is much more subtle. The best way to illustrate this is by means of what we call the memory-stick attack. Imagine a situation in which the provider of the devices is the adversary and has access to a proper RNG. The provider uses this RNG to generate a long sequence of good random numbers, stores them into a memory stick and sells it as a proper RNG to the user. The numbers generated by the user will pass any statistical test and look random; however, they are not properly random because they can be perfectly predicted by the adversary.

The generation of good randomness is a notoriously difficult problem^{2,3}. Basically, there are three types of approach. First, pseudo-random-number generators (PRNGs) use an algorithm to process an initial random seed. They are fast, cheap and the properties of the generated sequences are good enough for some applications. However, the random character of the output of PRNGs and their privacy are based on assumptions about the computational power of the adversary. But this is not the criterion adopted here—we demand unpredictability by any observer, independently of its computational power.

The second type of RNG is called a true RNG (TRNG) and exploits physical processes that are hard to predict, such as meteorological phenomena or the mouse movements of a computer user. Finally, there are quantum RNGs (QRNGs), which exploit a quantum process believed to be fundamentally random. In what follows, we focus the discussion on QRNGs, although many of the problems also apply to TRNGs.

The paradigmatic example of a QRNG is defined by the clicks observed after single photons impinge a beam-splitter; see Fig. 1. However, this idealized set-up may be difficult to implement in an experiment. More generally, imperfections on the devices are unavoidable and could deteriorate the quality of the generated numbers in uncontrolled manners. It is also difficult to exclude memory effects, for instance at the detectors, which produce correlations among the generated bits. The privacy of the symbols follows from the fact that the single-photon state is assumed to

¹ICFO-Institut de Ciències Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels, Barcelona, Spain. ²ICREA—Institució Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain. ³Department of Physics and Astronomy, University College of London, London WC1E 6BT, UK.

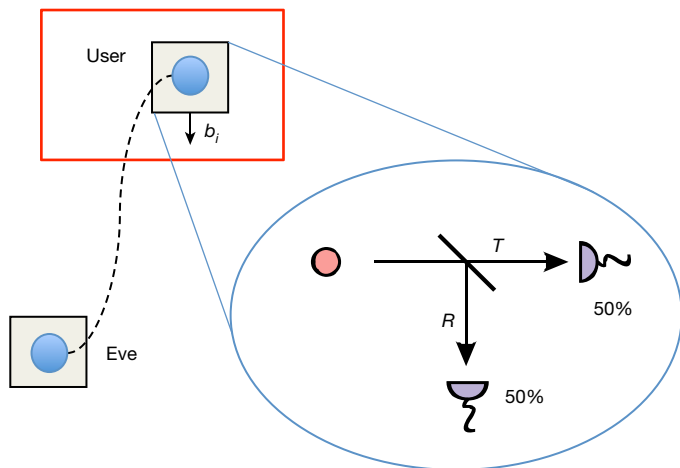


Figure 1 | Schemes for randomness generation. The user, in their secure location, has access to a device that generates bits b_i . The user wants to make sure that the value of these bits cannot be predicted by any observer outside their laboratory. The way to model this is by an external ‘super-observer’ who has access to any physical system other than that at the user’s location, which we represent by another device that may be correlated to the user’s. It is useful to interpret the external observer as an adversary or eavesdropper, ‘Eve’, who wants to predict the generated bits (for instance to hack any possible use for cryptographic applications). The generated bits should be unpredictable by Eve, even after measuring their device. For standard QRNGs, the random character of the outputs follows from assumptions about the inner working of the user’s device. The figure displays a scheme based on a single photon (red ball) impinging a beam-splitter with a transmission coefficient T , and reflection coefficient R , of $1/2$. Two single-photon detectors placed at the two arms of the interferometer measure the path taken by the photon. According to quantum physics, this process is probabilistic and the probability that a given detector ‘clicks’ is equal to the transmission coefficient of the beam-splitter, which is assumed to be $1/2$.

be pure and therefore cannot be coupled to another system. Hence, there are many assumptions about the working of the QRNG that are crucial to guarantee the perfect match between the ideal theoretical situation and the implementation, which is in turn necessary to guarantee the quality of the generated outputs.

The current certification method applied to QRNGs consists of passing statistical tests. Apart from the problems already mentioned, the use of these tests is even more problematic for QRNGs because they can also be satisfied by classical RNGs. So, the only guarantee the user has that the symbols have a quantum origin is the word of the provider. Furthermore, the user has no means to test the privacy of the symbols because they cannot rule out the memory-stick attack. Trusting the provider is again the only option. This level of trust between provider and user is unsatisfactory because in many situations, especially for cryptographic applications, it is convenient to reduce the trust on the provider as much as possible. And, even if the provider is trusted and has constructed the devices in the best possible way, uncontrolled drifts and changes on the devices are unavoidable and could deteriorate the quality of the generated randomness. Consequently, there is a need for solutions that certify the ‘quantumness’, quality and privacy of QRNGs without requiring any detailed modelling of the devices.

Device-independent QRNGs (DIQRNGs) offer a solution to the previous issues and provide protocols for generating certified randomness based on only general assumptions about the set-up, such as the validity of quantum physics. In particular, they do not require any assumptions about the inner working of the devices, which can be seen as quantum black boxes processing classical information. The development of DIQRNG protocols is an active research field involving many concepts and methods, including information-theoretical studies to design stronger protocols based on weaker assumptions and their experimental realization using current or near-future technology.

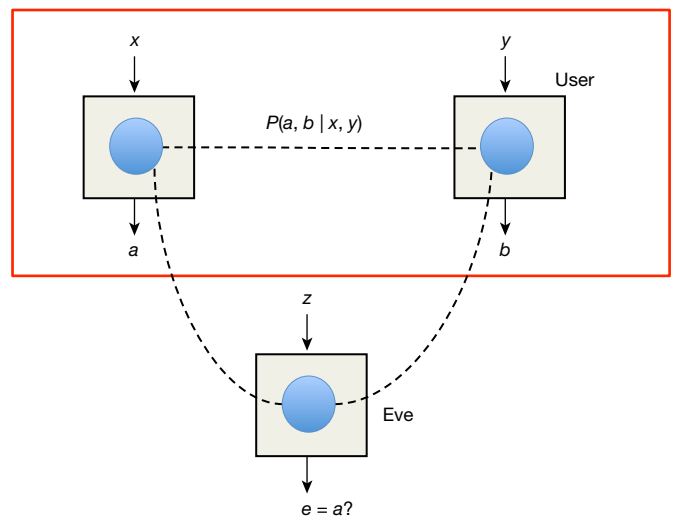


Figure 2 | Structure of DIQRNG protocols. In a general protocol for a DIQRNG, the user has access to $n \geq 2$ correlated devices. The figure shows the simplest case of two devices (boxes inside the red rectangle), which generate classical outputs a and b after applying the inputs x and y . The inputs x and y can be understood as the labels of the measurements performed on each device and the outputs as the obtained results. The external (eavesdropping) observer, Eve, may have a system that is correlated with the user’s devices. The randomness of one of the outputs, say a , can be quantified by the optimal probability P_{guess} that Eve guesses it correctly ($e = a$) after performing a measurement z on their system^{14,15,77}. In the case of quantum eavesdroppers, the guessing probability is optimized over all quantum preparations, including the tripartite state and measurements, which are compatible with the correlations observed by the user^{15,88–90}. The assumption of the validity of quantum mechanics can be relaxed by considering eavesdroppers who can prepare any tripartite correlations that are compatible with the no-signalling principle, even those that are incompatible with quantum physics^{15,90}.

In what follows, we first show how randomness certification without assumptions about the inner working of the devices can be achieved by exploiting the quantum violation of Bell inequalities. We then describe the state of the art in DIQRNG protocols and their experimental implementations. Because these implementations turn out to be challenging, we describe other approaches to certified randomness with milder experimental requirements. We also explain how, in addition to practical applications, protocols for certified randomness answer some fundamental questions in physics. We conclude with an outlook.

Device-independent randomness generation

DIQRNG protocols make use of the correlations observed when measuring entangled particles that do not have a classical analogue, as certified by the violation of a Bell inequality⁴. For such protocols, the user needs at least two separated devices to run a Bell test; see Fig. 2. The devices receive classical inputs x and y , and produce classical outputs a and b . After N_r rounds of collecting the data (x, y, a, b) , the user calculates the relative frequencies of the outcomes given the inputs $P(a, b | x, y)$, which can be estimated without making any assumptions about the internal working of the devices. A Bell inequality is a linear function of these relative frequencies

$$\beta = \sum c_{abxy} P(a, b | x, y) \leq \beta_L$$

characterized by the coefficients c_{abxy} . Here, β_L is the local bound satisfied by classical theories, such as that of Einstein–Podolsky–Rosen (EPR)⁵. The violation of the Bell inequality indicates the presence of non-classical correlations between the two devices.

The idea behind DIQRNGs is that, if the user observes a Bell inequality violation, they can be certain that the unknown quantum state in the devices is entangled and pure. The purity of the quantum state certifies

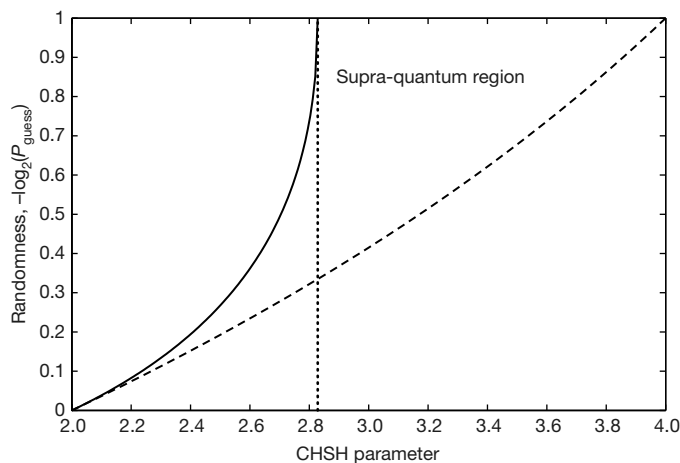


Figure 3 | Randomness as a function of the CHSH Bell inequality violation observed by the user. The amount of generated randomness at one of the outputs is shown, expressed as $-\log_2(P_{\text{guess}})$, where P_{guess} is the optimal guessing probability of the eavesdropper, such that a value of 0 corresponds to an adversary being able to correctly predict all outputs and a value of 1 to a perfectly random bit (for which the adversary will correctly guess the outputs half the time). The CHSH Bell inequality violation is expressed in terms of the CHSH parameter, with a value of 2 corresponding to no violation (the local bound). The curves are computed using the techniques in refs 15 and 77. The solid line refers to a quantum eavesdropper; the dashed line is for a non-signalling eavesdropper. The quantum violation of the CHSH inequality is upper-bounded by $2\sqrt{2}$ (vertical dotted line), but it is possible to get larger, ‘supra-quantum’ violations without breaking the no-signalling principle. At the local bound (at which the CHSH parameter is equal to 2), no randomness can be certified, whereas some randomness appears for any violation, that is, when the CHSH parameter is larger than 2. In both cases, a perfect random bit is certified by the corresponding maximal Bell violation, which corresponds to a CHSH parameter of $2\sqrt{2}$ in the quantum case and 4 for non-signalling correlations.

that the two devices are not too correlated with the environment or the external observer. The entanglement certifies that the local state of one of the devices is mixed and, thus, a measurement on it generates random outcomes. Moreover, the Bell certification of randomness is intrinsically quantum, because classical devices always satisfy a Bell inequality. Finally, the randomness is device-independent because only the observed statistics $P(a, b|x, y)$ are needed for its computation.

Consider a case in which the user tests the Clauser–Horne–Shimony–Holt (CHSH) Bell inequality⁶ and observes its maximal quantum violation¹. (Although some of the following results are illustrated using the CHSH inequality—the simplest Bell inequality—the main ideas and concepts apply to any Bell inequality.) According to quantum physics, the statistics of local measurements $\Pi_{a|x}$ and $\Pi_{b|y}$ on the state describing the two quantum devices $|\Psi\rangle$ are given by

$$P(a, b|x, y) = \langle \Psi | \Pi_{a|x} \otimes \Pi_{b|y} | \Psi \rangle \quad (2)$$

A maximal violation of the CHSH inequality is obtained only when projective measurements are performed on a maximally entangled state of two qubits^{7–9}

$$|\psi\rangle = |\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

Because the state is pure, it cannot be correlated to the environment. The local measurements on half of it produce perfect random bits according to equation (1), which are certified by the observed Bell violation. This intuitive argument holds when the maximal quantum violation of the CHSH inequality is obtained. For noisy, non-maximal violations, it is also possible to quantify the amount of randomness from the observed violation; see Fig. 3.

Although the previous discussion has exploited the properties of quantum correlations described by equation (2), it is possible to design DIQRNGs that do not rely on the validity of quantum theory, but on only that of the no-signalling principle, that is, the impossibility of faster-than-light communication between devices. Under the sole assumption of no-signalling, the violation of a Bell inequality guarantees the random character of the outputs^{10–16}.

All these nice features come at a price: the user should meet the conditions needed in equation (2). They must ensure that:

- (C1) the inputs (x, y) have no correlations with the devices; and
- (C2) there is no communication between the two devices during the generation of the two distant outcomes (see Fig. 2).

Looking at equation (2), condition (C1) implies, for example, that the measured quantum state is independent of x and y . Condition (C2) imposes the tensor product and that the measurements on one device do not depend on the input of the other. To guarantee these conditions, the user needs to make physical assumptions about the devices (albeit not about their internal working).

Protocols

We now provide a unified description of most of the protocols for DIQRNGs that have been proposed so far. In general, these protocols involve $n \geq 2$ devices, each having an input x_i and an output a_i for $i = 1, \dots, n$. Condition (C1) imposes that the inputs (x_1, \dots, x_n) must be selected in a way that is uncorrelated to the devices. A standard (but not the only) way of satisfying this condition is by choosing the inputs using a random seed, which has to meet certain ‘independence’ requirements depending on the protocol (see Box 1).

A series of works have focused on efficiency, trying to optimize the trade-off between initial and final randomness. The corresponding protocols are known as randomness expansion protocols^{14,15,17–23}. It has been proven that an unbounded amount of randomness (length of the final random bit string $N_k \rightarrow \infty$) can be generated from a finite seed^{19,21}.

Other works^{19,21,24–29} have focused on the quality of the seed and studied how arbitrarily good randomness can be generated in Bell set-ups using sources of imperfect randomness. These protocols are often known as ‘randomness amplification’ protocols²⁴. A commonly used model for the imperfect seed (s_1, s_2, \dots) is a Santha–Vazirani source³⁰, which is characterized by a parameter $\epsilon \in (0, 1/2)$ such that

$$\epsilon \leq P(s_i | s_1, s_2, \dots, s_{i-1}, \text{devices, Eve}) \leq 1 - \epsilon \quad (3)$$

for all i , in which ‘Eve’ denotes any correlated information owned by the eavesdropper. Therefore, the larger the value of ϵ , the higher the randomness of the bits. Other models for the seed, more general than the Santha–Vazirani source, have also been considered, such as min-entropy sources^{21,27}. In the case of Santha–Vazirani (equation (3)), the performance of randomness amplification protocols is measured by comparing the parameter ϵ_i of the initial source with the final ϵ_f of the generated bits. Full randomness amplification is attained when a source with $\epsilon_i \rightarrow 0$ is mapped to one with $\epsilon_f \rightarrow 1/2$ (ref. 29). From a fundamental point of view, the existence of protocols that attain full randomness amplification is important; we return to this point below. However, from a practical point of view, although allowing for a non-perfect seed is a good addition to the security of a protocol, in most applications one can assume that the seed is uncorrelated to the adversary and devices, in which case the expansion rate is a more practical figure of merit. Randomness expansion and amplification protocols are asides of the general problem, which is the generation of device-independent private randomness under the minimal set of assumptions and with minimal resources (see Table 1).

Bell-certified randomness is also a resource for device-independent quantum key distribution^{11,13,23,31–42}. However, the goal here is not only to generate randomness, but to establish a secret key between two distant users using a Bell violation observed between their devices. In particular, and in contrast to the case of randomness generation, the devices are held

BOX 1

DIQRNG PROTOCOL

Steps of the protocol

Data collection. Repeat steps (a), (b), (c) N_r times:

(a) A source of n -partite entangled states sends a particle to each of the n devices.

(b) Part of the seed is processed to generate a sample from the prior distribution $P(x_1, \dots, x_n)$ of the inputs applied to each device. This distribution can be optimized before implementing the protocol to suit the statistics of the devices.

(c) Measurement x_i is performed on device i generating outcome a_i . The inputs and outputs of the devices ($a_1, \dots, a_n, x_1, \dots, x_n$) are stored.

Non-locality estimation. Calculate the relative frequency of every combination of inputs and outputs $P_{\text{freq}}(a_1, \dots, a_n | x_1, \dots, x_n)$ using the raw data collected in the N_r rounds. From this data, estimate the non-locality of the observed correlations and determine the length N_k of the final random bit string. The larger the amount of non-locality, the longer N_k . If the observed non-locality is insufficient, then the protocol is aborted ($N_k = 0$).

Classical post-processing. Generate the final N_k -bit string using the raw data collected in the N_r rounds plus an additional part of the seed. This process is often made with a randomness ‘extractor’^{91,92}.

Parameters of the protocol

The parameters that are relevant to the design of DIQRNG protocols are as follows.

Efficiency. This parameter describes the trade-off between the amount of randomness generated and the resources consumed by the DIQRNG protocol. Examples of these resources are the number of random bits of the seed N_s or the number of times the devices are used N_r .

Quality of the seed. The random seed may not be perfect. For example, the seed may not necessarily be uniformly distributed, or display correlations with the devices or the adversary. Instead of using a seed, another possibility is to assume the existence of a good and free public source of randomness, such as the broadcast by the National Institute of Standards and Technology (NIST) Randomness Beacon (http://www.nist.gov/itl/csd/ct/nist_beacon.cfm). In this case, the protocol generates private randomness from public randomness.

Robustness. This parameter characterizes the tolerance of the protocol to noise and imperfections. This allows for using realistic noisy apparatuses. A protocol is robust if it works ($N_k > 0$) for violations above some threshold, which does not need to coincide with the local bound.

Number of devices. The minimum number of devices used in the protocol is $n = 2$.

Composability. Protocols should be such that, if the adversary learns some information about the N_k final random bits, then they should be able to deduce essentially no additional information⁹³.

Physical assumptions. Many protocols assume that all the devices and the adversary are constrained by the laws of quantum mechanics. However, it is possible to relax this requirement and assume only the validity of the no-signalling principle. Even when the security of the protocol does not rely on the validity of quantum mechanics, quantum technology is still needed to generate Bell-violating correlations.

in two separate locations and the channel between them is accessible to the eavesdropper.

Implementations

The implementation of the aforementioned DIQRNG protocols requires the observation of a Bell inequality violation. For that, it is necessary to prepare an entangled state of $n \geq 2$ particles, which are distributed to n devices where they are subjected to local measurements. Assuming the validity of quantum physics, the experimental set-up should guarantee that conditions (C1) and (C2) are met so that the observed statistics are correctly described by equation (2).

An experimental challenge for the observation of Bell inequality violations is that a high detection efficiency (more than about 70%) is required to close the detection loophole⁴³. This loophole implies that, for low enough detection efficiencies, the statistics of a Bell experiment can always be described by an EPR model, which is deterministic, and so no randomness certification is possible^{44,45}. Closing the detection loophole is demanding because it concerns any losses in the set-up.

Table 1 | State-of-the-art DIQRNG protocols

	n	$N_k(N_s)$	N_k/N_r	Seed quality	Seed privacy	QM
Ref. 21	Large	Infinite	0	$\epsilon > 0$	No	Yes
Ref. 20	2	Exponential	> 0	$\epsilon = 1/2$	Yes	Yes
Ref. 28	2	Small	0	$\epsilon > 0$	Yes	No
Ideal	2	Infinite	> 0	$\epsilon > 0$	No	No

Properties of the best known protocols, which are robust, composable and immune to attacks that exploit memory effects on the devices, as in the memory loophole introduced in the context of Bell inequality violations⁸⁷. The parameters are: the number of devices n ; the amount of expansion of the initial seed $N_k(N_s)$, expressed as the way in which the length of the final random bit string N_k depends on that of the initial seed N_s ; the efficiency rate N_k/N_r , where N_r is the number of times the devices are used; the seed quality ϵ in terms of equation (3); whether publishing the seed compromises security; and whether the security of the protocol relies on the validity of quantum mechanics (QM). Note that infinite $N_k(N_s)$ assumes that N_s is sufficiently long. The bottom row contains the optimal value for each parameter.

But it is a loophole that can be, and actually has been^{46–53}, completely closed, because it does not put into question the validity of the description given in equation (2), but simply demands a high enough detection efficiency.

Contrary to the detection loophole, the locality⁵⁴, collapse-locality⁵⁵ and free-will⁵⁶ loopholes do put into question the validity of equation (2). Consequently, they can never be strictly closed, but their plausibility can be enforced by making physically motivated assumptions about the experimental arrangement. The locality loophole affects condition (C2), that is, whether the measurements in either device depend on what happens on the other device. The standard approach to closing this loophole is based on Einstein’s theory of relativity and involves arranging the measurements so that they define events that are separated in a space-like way such that no communication can take place between the two devices. This is a very satisfactory solution, but also a demanding one. In our view, there are relevant scenarios in which it is also possible to assume the validity of (C2) without space-like-separated measurements. For example, a low level of trust may be put on the provider so that it is safe to assume that the devices do not signal to each other when producing the outputs given the inputs. Or some level of shielding, essential for any cryptographic use of the generated numbers, may be assumed, which can in turn be used to avoid any unwanted communication between the devices.

Moreover, the space-like arrangement that is usually adopted to close the locality loophole also assumes that there is precise knowledge about when the local measurements start and end, that is, when the inputs x, y are defined and the outputs a, b are produced. This issue relates to the free-will and collapse-locality loopholes, which also put into question condition (C1). If there is no timing information about when the inputs are generated, then this information could, for instance, exist before the entangled state is produced; that is, the state in equation (2) could depend on x and y . A proposed theoretical way to avoid this dependence is to assume that the inputs are generated by humans, hence the term ‘free will’. The usual and more practical approach to close the loophole is to use a standard QRNG⁵⁶. However, it is questionable whether (and if so, why) a QRNG is preferable over other processes of classical origin⁵⁷. Similar considerations apply to the timing of the outputs, as in the collapse-locality loophole: one needs to define when the classical results are actually produced to guarantee that the measurements define space-like-separated events.

Taking into account all of these points and the technological state of the art, two set-ups emerge in which to implement DIQRNG protocols: entangled particles in separate locations and entangled photons. In the first case, a Bell test is performed between two distant massive particles, such as nitrogen-vacancy centres⁵¹, or ions in two traps^{47,48} that have been entangled through entanglement swapping on two photon–particle entangled pairs^{58,59}. The advantage of this set-up is that massive particles can be measured with almost perfect efficiency, thus closing the detection loophole. A first proof-of-principle demonstration of a DIQRNG, reporting a generation rate of 42 random bits after approximately one month of measurements, was performed using two entangled ions in two traps at 1-m distance¹⁵. This rate is valid under the assumption that the experimental set-up was not operating in a malicious way¹⁷. Although

challenging, set-ups with entangled distant particles also allow the measurements to be arranged so that one can reasonably assume that both the detection and locality loopholes are closed. This was achieved in ref. 51 and subsequently in ref. 60; however, these experiments do not report any analysis of random-number generation.

In the second case, the set-up consists of performing polarization measurements on two entangled photons. Historically, one of the main challenges in these set-ups was that photon-detection efficiencies were too low to close the detection loophole; however, advances in photo-detectors have allowed it to be closed^{49,50}. The first experiment focused on only the Bell violation, but the second reported a random-number generation rate of 0.4 bits s^{-1} . More recently, the locality loophole has also been closed in photonic experiments^{52,53}; but again, none of these experiments was analysed in the context of device-independent random-number generation.

Other methods for randomness generation

Because meeting conditions (C1) and (C2) in an experiment is challenging, alternative proposals for certified quantum randomness generation have been proposed. The idea is to keep part of the device-independent spirit and make only mild assumptions about the set-up, yet without any detailed modelling of the devices. Randomness certification comes from a purely quantum effect with no classical analogue, under the mentioned assumptions. Standard QRNGs do not fit into this category because they require modelling and certify randomness using statistical tests that are also satisfied by classical RNGs.

A series of works have explored information protocols under a dimensional constraint^{61,62}—a scenario known as semi-device-independent. The set-up is different from a Bell test and consists of a preparing device that prepares a system in different quantum states and a measuring device that performs measurements on it. It is assumed that the states prepared by the first device and measured by the second belong to a Hilbert space of dimension not larger than d ; this is the extra assumption that goes beyond the fully device-independent paradigm. Randomness certification is then obtained via the violation of the so-called dimension witnesses⁶³. One of the practical advantages of this approach is that it does not require the generation of entanglement. The required detection efficiencies are smaller, but still demanding to achieve experimentally⁴⁵. A solution to this problem was suggested in ref. 64, where it was shown that schemes in which one assumes that the preparation and measuring device share no correlations, and that devices do not display memory effects, certify the presence of randomness for any value of the detection efficiency. A proof-of-principle experimental demonstration of this proposal has been performed⁶⁵, but a security proof for these schemes without assumptions about memory effects is lacking.

A second proposal considers an asymmetric scenario in which some of the devices are fully trusted; for example, one may trust the preparing devices, but not the measuring devices. Asymmetric scenarios are often considered in the context of steering⁶⁶. This concept is defined in the same set-up as non-locality, in which two parties perform measurements on two distant quantum particles, but now one of the devices is fully trusted. The detection of steering provides a quantum certification that is sufficient to guarantee the presence of randomness⁶⁷. Steering has been experimentally demonstrated with the detection and locality loopholes closed^{68–70}, but the detection efficiencies needed for randomness expansion still pose a challenge⁴⁵. Security proofs are in a preliminary stage also in the case of steering. A general security proof was provided in ref. 71, but it requires a very low level of noise.

Fundamental questions on randomness

The results connecting randomness and non-locality are relevant not only for applications of quantum technologies, but also for our understanding of physics. Protocols that attain full randomness amplification against non-signalling eavesdroppers represent the strongest form of certification using quantum physics of the existence of random events in nature. It is impossible to certify randomness from scratch. Under only

the assumption of no-signalling, the violation of Bell inequalities certifies the presence of randomness, but also requires some initial randomness. Full randomness amplification protocols^{19,21,25–29} are not able to completely break this circularity, but relax it as much as possible.

Historically, the discussion on EPR models and Bell inequalities was motivated by the search for a ‘complete’ alternative to quantum theory, in the sense that measurement outcomes could have a deterministic description within the alternative theory. The violation of Bell inequalities implies that quantum predictions cannot be reproduced by a deterministic theory without violating the no-signalling principle. In recent years, stronger proofs of the ‘uncompleteness’ of quantum theory have appeared^{11,13,72–74}. These works show that a no-signalling model, possibly non-deterministic, that has higher predictive power than quantum theory does not exist. Full randomness amplification protocols against no-signalling eavesdroppers can also be seen as proofs of the uncompleteness of quantum theory. The study of randomness is also relevant when comparing quantum theory with more general theories that respect the no-signalling principle^{9,75}. It has been shown that theories leading to general non-signalling correlations do not allow for maximal randomness certification, whereas quantum theory does⁷⁶.

Finally, a series of works have shown that the relation between entanglement, non-locality and randomness is more subtle than expected. For instance, states with arbitrarily small amounts of entanglement (and non-locality) allow for maximal randomness certification⁷⁷. Recent progress in this direction shows that the use of more complex measurements, such as non-projective⁷⁸ or sequences of measurements⁷⁹, provides further advantages for randomness certification. Despite these results, a complete understanding of the relation between entanglement, non-locality and randomness is still missing.

Outlook

DIQRNG protocols address fundamental and practical drawbacks of standard RNG schemes and so represent a change of paradigm for randomness generation. On the theory side, the existing security proofs show the validity of the approach. However, further theoretical studies are needed to understand how to relax the requirements for DIQRNGs. The ultimate goal would be to design a robust and composable secure protocol that attains an infinite randomness expansion rate using initial sources of arbitrarily weak public randomness with only two devices, and assuming only the validity of the no-signalling principle (see Table 1). Although this ambitious goal may be unreachable, there is still a lot of room for improvement on the conditions that are needed for device-independent randomness generation.

On the implementation side, it is expected that new DIQRNG experiments using the set-ups explained above will be reported in the coming years, with constantly improving generation rates. Looking ahead, integrated photonic circuits and solid-state set-ups present as other platforms in which to run the previous protocols. To our knowledge, however, no Bell experiment has been reported on integrated photonic circuits. Non-local correlations in solid-state devices, as certified by the violation of a Bell inequality, have been reported for two-^{80,81} and three-component systems^{82,83}. These technologies could be more promising in terms of miniaturization and, hence, the possible construction of commercial DIQRNGs. However, miniaturization comes at a price: the possible validity of condition (C2) becomes less clear, even if one accepts that measurements do not need to be space-like separated. Theoretical solutions to take into account cross-talk effects have been proposed⁸⁴. In fact, the analysis of implementations of DIQRNGs leads to new theoretical questions, such as being able to determine: (i) the detection efficiencies required for randomness expansion⁴⁵; (ii) which Bell set-ups are more robust against noise or detection inefficiencies⁸⁵; and (iii) the ways in which to deal with detection inefficiencies⁸⁶.

Continued advances in the theory and implementation of DIQRNGs will lead to the design of more robust and feasible schemes. RNGs with unprecedented standards of quality and security seem within reach using quantum technologies.

Received 17 May; accepted 23 September 2016.

1. Rukhin, A. *et al.* *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*. NIST Special Publication 800-22rev1a, <http://csrc.nist.gov/groups/ST/toolkit/rng/index.html> (National Institute of Standards and Technology, 2010).
2. von Neumann, J. Various techniques used in connection with random digits. *Natl Bureau Stand. Appl. Math. Ser.* **12**, 36–38 (1951).
3. Markowsky, G. The sad history of random bits. *J. Cyber Secur. Mobil.* **3**, 1–24 (2014).
4. Bell, J. S. On the Einstein Podolsky Rosen paradox. *Physics* **1**, 195–200 (1964).
5. Einstein, A., Podolsky, B. & Rosen, N. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, 777–780 (1935).
6. Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880–884 (1969).
7. Khalfin, L. A. & Tsirelson, B. S. in *Symposium on the Foundations of Modern Physics* (eds Lahti, P. & Mittelstaedt, P.) 441–460 (World Scientific, 1985).
8. Tsirelson, B. S. Some results and problems on quantum Bell-type inequalities. *Hadronic J.* **8** (Suppl.), 329–345 (1993).
9. Popescu, S. & Rohrlich, D. Quantum nonlocality as an axiom. *Found. Phys.* **24**, 379–385 (1994).
10. Valentini, A. Signal-locality in hidden-variables theories. *Phys. Lett. A* **297**, 273–278 (2002).
11. Barrett, J., Hardy, L. & Kent, A. No signaling and quantum key distribution. *Phys. Rev. Lett.* **95**, 010503 (2005).
A quantum key distribution protocol is described that is secure against non-signalling eavesdroppers on the basis of Bell inequality violations.
12. Masanes, L., Acín, A. & Gisin, N. General properties of nonsignaling theories. *Phys. Rev. A* **73**, 012112 (2006).
13. Barrett, J., Kent, A. & Pironio, S. Maximally non-local and monogamous quantum correlations. *Phys. Rev. Lett.* **97**, 170409 (2006).
14. Colbeck, R. *Quantum and Relativistic Protocols for Secure Multi-party Computation*. PhD Thesis, Univ. Cambridge (2006); preprint at <http://arxiv.org/abs/0911.3814>.
Chapter 5 of this thesis introduces the idea of private randomness generation certified by Bell inequality violations.
15. Pironio, S. *et al.* Random numbers certified by Bell's theorem. *Nature* **464**, 1021–1024 (2010).
This paper presents a general construction of device-independent randomness expansion protocols and an experimental demonstration thereof.
16. Hall, M. J. W. Complementary contributions of indeterminism and signalling to quantum correlations. *Phys. Rev. A* **82**, 062117 (2010).
17. Colbeck, R. & Kent, A. Private randomness expansion with untrusted devices. *J. Phys. A* **44**, 095305 (2011).
18. Vazirani, U. V. & Vidick, T. Certifiable quantum dice. *Phil. Trans. R. Soc. Lond. A* **370**, 3432–3448 (2012).
19. Coudron, M. & Yuen, H. Infinite randomness expansion and amplification with a constant number of devices. In *Proc. 46th Annual ACM Symposium on Theory of Computing* 427–436 (ACM New York, 2014).
20. Miller, C. A. & Shi, Y. Robust protocols for securely expanding randomness and distributing keys using untrusted quantum devices. In *Proc. 46th Annual ACM Symposium on Theory of Computing* 417–426 (ACM New York, 2014).
21. Chung, K. M., Shi, Y. & Wu, X. Physical randomness extractors: generating random numbers with minimal assumptions. Preprint at <http://arxiv.org/abs/1402.4797> (2015).
22. Miller, C. A. & Shi, Y. Universal security for randomness expansion from the spot-checking protocol. Preprint at <http://arxiv.org/abs/1411.6608> (2015).
23. Arnon-Friedman, R., Renner, R. & Vidick, T. Simple and tight device-independent security proofs. Preprint at <http://arxiv.org/abs/1607.01797> (2016).
24. Colbeck, R. & Renner, R. Free randomness can be amplified. *Nat. Phys.* **8**, 450–454 (2012).
25. Gallego, R. *et al.* Full randomness from arbitrarily deterministic events. *Nat. Commun.* **4**, 2654 (2013).
26. Brandão, F. G. S. L. *et al.* Robust device-independent randomness amplification with few devices. *Nat. Commun.* **7**, 11345 (2016).
27. Bouda, J., Pawłowski, M., Pivoluska, M. & Plesch, M. Device-independent randomness extraction for arbitrarily weak min-entropy source. *Phys. Rev. A* **90**, 032313 (2014).
28. Ramanathan, R. *et al.* Randomness amplification against no-signaling adversaries using two devices. Preprint at <http://arxiv.org/abs/1504.06313> (2015).
29. Wojewódka, H. *et al.* Amplifying the randomness of weak sources correlated with devices. Preprint at <http://arxiv.org/abs/1601.06455> (2016).
30. Santha, M. & Vazirani, U. Generating quasi-random sequences from slightly-random sources. In *Proc. 25th IEEE Symposium on Foundations of Computer Science* 434–440 (IEEE, 1984).
31. Mayers, D. & Yao, A. Self testing quantum apparatus. *Quantum Inf. Comput.* **4**, 273–286 (2004).
This work introduces the idea of quantum information protocols with untrusted devices.
32. Acín, A., Gisin, N. & Masanes, L. From Bell's theorem to secure quantum key distribution. *Phys. Rev. Lett.* **97**, 120405 (2006).
33. Acín, A. *et al.* Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.* **98**, 230501 (2007).
34. Masanes, L. Universally-composable privacy amplification from causality constraints. *Phys. Rev. Lett.* **102**, 140501 (2009).
35. Pironio, S. *et al.* Device-independent quantum key distribution secure against collective attacks. *New J. Phys.* **11**, 045021 (2009).
36. Hänggi, E. & Renner, R. Device-independent quantum key distribution with commuting measurements. Preprint at <http://arxiv.org/abs/1009.1833> (2010).
37. Masanes, L., Pironio, S. & Acín, A. Secure device-independent quantum key distribution with causally independent measurement devices. *Nat. Commun.* **2**, 238 (2011).
38. Masanes, L., Renner, R., Christandl, M., Winter, A. & Barrett, J. Full security of key distribution from no-signaling constraints. *IEEE Trans. Inf. Theory* **60**, 4973–4986 (2014).
39. Barrett, J., Colbeck, R. & Kent, A. Unconditionally secure device-independent quantum key distribution with only two devices. *Phys. Rev. A* **86**, 062326 (2012).
40. Reichardt, B., Unger, F. & Vazirani, U. Classical command of quantum systems. *Nature* **496**, 456–460 (2013).
41. Barrett, J., Colbeck, R. & Kent, A. Memory attacks on device-independent quantum cryptography. *Phys. Rev. Lett.* **110**, 010503 (2013).
42. Vazirani, U. & Vidick, T. Fully device independent quantum key distribution. *Phys. Rev. Lett.* **113**, 140501 (2014).
43. Pearle, P. Hidden-variable example based upon data rejection. *Phys. Rev. D* **2**, 1418–1425 (1970).
44. Massar, S. & Pironio, S. Violation of local realism vs detection efficiency. *Phys. Rev. A* **68**, 062109 (2003).
45. Acín, A., Cavalcanti, D., Passaro, E., Pironio, S. & Skrzypczyk, P. Necessary detection efficiencies for secure quantum key distribution and bound randomness. *Phys. Rev. A* **93**, 012319 (2016).
46. Rowe, M. A. *et al.* Experimental violation of a Bell's inequality with efficient detection. *Nature* **409**, 791–794 (2001).
47. Matsukevich, D. N., Maunz, P., Moehring, D. L., Olmschenk, S. & Monroe, C. Bell inequality violation with two remote atomic qubits. *Phys. Rev. Lett.* **100**, 150404 (2008).
48. Hofmann, J. *et al.* Heralded entanglement between widely separated atoms. *Science* **337**, 72–75 (2012).
49. Giustina, M. *et al.* Bell violation with entangled photons, free of the fair-sampling assumption. *Nature* **497**, 227–230 (2013).
50. Christensen, B. G. *et al.* Detection-loophole-free test of quantum nonlocality, and applications. *Phys. Rev. Lett.* **111**, 130406 (2013).
This paper provides an experimental demonstration of device-independent randomness expansion using entangled photons.
51. Hensen, B. *et al.* Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **526**, 682–686 (2015).
An experimental Bell violation, free of the detection and locality loopholes, is demonstrated using entangled nitrogen-vacancy centres.
52. Giustina, M. *et al.* Significant-loophole-free test of Bell's theorem with entangled photons. *Phys. Rev. Lett.* **115**, 250401 (2015).
An experimental Bell violation, free of the detection and locality loopholes, is demonstrated using entangled photons.
53. Shalm, L. K. *et al.* A strong loophole-free test of local realism. *Phys. Rev. Lett.* **115**, 250402 (2015).
An experimental Bell violation, free of the detection and locality loopholes, is demonstrated using entangled photons.
54. Aspect, A. Proposed experiment to test separable hidden-variable theories. *Phys. Lett. A* **54**, 117–118 (1975).
55. Kent, A. Causal quantum theory and the collapse locality loophole. *Phys. Rev. A* **72**, 012107 (2005).
56. Scheidl, T. *et al.* Violation of local realism with freedom of choice. *Proc. Natl Acad. Sci. USA* **107**, 19708–19713 (2010).
57. Pironio, S. Random 'choices' and the locality loophole. Preprint at <http://arxiv.org/abs/1510.00248> (2015).
58. Żukowski, M., Zeilinger, A., Horne, M. A. & Ekert, A. K. "Event-ready-detectors" Bell experiment via entanglement swapping. *Phys. Rev. Lett.* **71**, 4287–4290 (1993).
59. Simon, C. & Irvine, W. T. M. Robust long-distance entanglement and a loophole-free Bell test with ions and photons. *Phys. Rev. Lett.* **91**, 110405 (2003).
60. Hensen, B. *et al.* Loophole-free Bell test using electron spins in diamond: second experiment and additional analysis. *Sci. Rep.* **6**, 30289 (2016).
61. Pawłowski, M. & Brunner, N. Semi-device-independent security of one-way quantum key distribution. *Phys. Rev. A* **84**, 010302(R) (2011).
62. Li, H.-W., Pawłowski, M., Yin, Z.-Q., Guo, G.-C. & Han, Z.-F. Semi-device-independent randomness certification using $n \rightarrow 1$ quantum random access codes. *Phys. Rev. A* **85**, 052308 (2012).
63. Gallego, R., Brunner, N., Hadley, C. & Acín, A. Device-independent tests of classical and quantum dimensions. *Phys. Rev. Lett.* **105**, 230501 (2010).
64. Bowles, J., Quintino, M. T. & Brunner, N. Certifying the dimension of classical and quantum systems in a prepare-and-measure scenario with independent devices. *Phys. Rev. Lett.* **112**, 140407 (2014).
65. Lunghi, T. *et al.* A self-testing quantum random number generator. *Phys. Rev. Lett.* **114**, 150501 (2015).
66. Wiseman, H. M., Jones, S. J. & Doherty, A. C. Steering, entanglement, nonlocality, and the EPR paradox. *Phys. Rev. Lett.* **98**, 140402 (2007).
67. Branciard, C., Cavalcanti, E. G., Walborn, S. P., Scarani, V. & Wiseman, H. M. One-sided device-independent quantum key distribution: security, feasibility, and the connection with steering. *Phys. Rev. A* **85**, 010301(R) (2012).

68. Smith, D. H. *et al.* Conclusive quantum steering with superconducting transition edge sensors. *Nat. Commun.* **3**, 625 (2012).
69. Bennet, A. J. *et al.* Arbitrarily loss-tolerant Einstein–Podolsky–Rosen steering allowing a demonstration over 1 km of optical fiber with no detection loophole. *Phys. Rev. X* **2**, 031003 (2012).
70. Wittmann, B. *et al.* Loophole-free Einstein–Podolsky–Rosen experiment via quantum steering. *New J. Phys.* **14**, 053030 (2012).
71. Tomamichel, M., Fehr, S., Kaniewski, J. & Wehner, S. A monogamy-of-entanglement game with applications to device-independent quantum cryptography. *New J. Phys.* **15**, 103002 (2013).
72. Colbeck, R. & Renner, R. No extension of quantum theory can have improved predictive power. *Nat. Commun.* **2**, 411 (2011).
73. Colbeck, R. & Renner, R. Is a system's wave function in one-to-one correspondence with its elements of reality? *Phys. Rev. Lett.* **108**, 150402 (2012).
74. Dhara, C., de la Torre, G. & Acín, A. Can observed randomness be certified to be fully intrinsic? *Phys. Rev. Lett.* **112**, 100402 (2014).
75. Barrett, J. *et al.* Non-local correlations as an information theoretic resource. *Phys. Rev. A* **71**, 022101 (2005).
76. de la Torre, G., Hoban, M. J., Dhara, C., Pretico, G. & Acín, A. Maximally nonlocal theories cannot be maximally random. *Phys. Rev. Lett.* **114**, 160502 (2015).
77. Acín, A., Massar, S. & Pironio, S. Randomness versus nonlocality and entanglement. *Phys. Rev. Lett.* **108**, 100402 (2012).
78. Acín, A., Pironio, S., Vértesi, T. & Wittek, P. Optimal randomness certification from one entangled bit. *Phys. Rev. A* **93**, 040102 (2016).
79. Curchod, F. J. *et al.* Unbounded randomness certification using sequences of measurements. Preprint at <http://arxiv.org/abs/1510.03394> (2015).
80. Ansmann, M. *et al.* Violation of Bell's inequality in Josephson phase qubits. *Nature* **461**, 504–506 (2009).
81. Dehollain, J. P. *et al.* Bell's inequality violation with spins in silicon. *Nat. Nanotechnol.* **11**, 242–246 (2016).
82. Neeley, M. *et al.* Generation of three-qubit entangled states using superconducting phase qubits. *Nature* **467**, 570–573 (2010).
83. DiCarlo, L. Preparation and measurement of three-qubit entanglement in a superconducting circuit. *Nature* **467**, 574–578 (2010).
84. Silman, J., Pironio, S. & Massar, S. Device-independent randomness generation in the presence of weak cross-talk. *Phys. Rev. Lett.* **110**, 100504 (2013).
85. Máttar, A., Skrzypczyk, P., Brask, J. B., Cavalcanti, D. & Acín, A. Optimal randomness generation from optical Bell experiments. *New J. Phys.* **17**, 022003 (2015).
86. Thinh, L. P., de la Torre, G., Bancal, J.-D., Pironio, S. & Scarani, V. Randomness in post-selected events. *New J. Phys.* **18**, 035007 (2016).
87. Barrett, B., Collins, D., Hardy, L., Kent, A. & Popescu, S. Quantum nonlocality, Bell inequalities and the memory loophole. *Phys. Rev. A* **66**, 042111 (2002).
88. Navascués, M., Pironio, S. & Acín, A. Bounding the set of quantum correlations. *Phys. Rev. Lett.* **98**, 010401 (2007).
89. Navascués, M., Pironio, S. & Acín, A. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New J. Phys.* **10**, 073013 (2008).
90. Nieto-Silleras, O., Pironio, S. & Silman, J. Using complete measurement statistics for optimal device-independent randomness evaluation. *New J. Phys.* **16**, 013035 (2014).
91. Trevisan, L. Extractors and pseudorandom generators. *J. Assoc. Comput. Mach.* **48**, 860–879 (2001).
92. Nisan, N. & Ta-Shma, A. Extracting randomness: a survey and new constructions. *J. Comput. Syst. Sci.* **58**, 148–173 (1999).
93. Canetti, R. Universally composable security: a new paradigm for cryptographic protocols. In *Proc. 42nd IEEE Symposium on Foundations of Computer Science* 136–145 (IEEE, 2001).

Acknowledgements We thank S. Pironio, V. Scarani, F. G. S. L. Brandao, R. Arnon Friedman and A. Wallraff for discussions. A.A. acknowledges support from the ERC CoG QITBOX, the AXA Chair in Quantum Information Science, Spanish MINECO (FOQUS FIS2013-46768-P and SEV-2015-0522), Fundación Cellex, the Generalitat de Catalunya (SGR 875) and The John Templeton Foundation. L.M. is supported by EPSRC.

Author Contributions Both authors equally contributed to the work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.A. (antonio.acin@icfo.es).

Safeguarding pollinators and their values to human well-being

Simon G. Potts¹, Vera Imperatriz-Fonseca², Hien T. Ngo³, Marcelo A. Aizen⁴, Jacobus C. Biesmeijer^{5,6}, Thomas D. Breeze¹, Lynn V. Dicks⁷, Lucas A. Garibaldi⁸, Rosemary Hill⁹, Josef Settele^{10,11} & Adam J. Vanbergen¹²

Wild and managed pollinators provide a wide range of benefits to society in terms of contributions to food security, farmer and beekeeper livelihoods, social and cultural values, as well as the maintenance of wider biodiversity and ecosystem stability. Pollinators face numerous threats, including changes in land-use and management intensity, climate change, pesticides and genetically modified crops, pollinator management and pathogens, and invasive alien species. There are well-documented declines in some wild and managed pollinators in several regions of the world. However, many effective policy and management responses can be implemented to safeguard pollinators and sustain pollination services.

Pollinators are inextricably linked to human well-being through the maintenance of ecosystem health and function, wild plant reproduction, crop production and food security. Pollination, the transfer of pollen between the male and female parts of flowers that enables fertilization and reproduction, can be achieved by wind and water, but the majority of the global cultivated and wild plants depend on pollination by animals. Although most animal pollinators are insects (for example, bees, flies, butterflies, moths, wasps, beetles and thrips), some vertebrate pollinators exist (for example, birds, bats and other mammals and lizards). Bees are the most important group of pollinators, visiting more than 90% of the leading 107 global crop types¹. Over 20,000 bee species have been described worldwide², of which up to 50 species are managed, and about 12 are commonly used for crop pollination, such as the western honeybee (*Apis mellifera*), the eastern honeybee (*Apis cerana*), some bumblebees, stingless bees and solitary bees. *Apis mellifera* is the most commonly managed bee in the world, although there is growing evidence highlighting the roles of wild pollinators and of diverse pollinator assemblages in contributing to global crop production³.

Our knowledge and response actions have not kept pace with the threats to pollinators and pollination services. Although there has been increased interest from science⁴, policymakers^{5,6} and the public, a mismatch remains between scientific evidence of impacts and conservation, and management responses. As a step towards further outreach to a wider audience, here we review the diverse values of pollinators, their status and trends, risks from environmental pressures and consequent management and policy response options, and highlight key knowledge gaps. Our review is robustly underpinned by the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services *Pollinators, Pollination and Food Production* assessment⁵, whose 77 international experts critically evaluated the available global evidence up until May 2015; we have also drawn upon key publications arising after this date.

Diversity of values of pollinators and pollination

Pollinators provide numerous benefits to humans, such as securing a reliable and diverse seed and fruit supply, sustaining populations of wild plants that underpin biodiversity and ecosystem function, producing honey and other beekeeping products, and supporting cultural values. Much of the recent international focus on pollination services has been on the benefits to food production. Animal pollination directly affects the yield and/or quality of approximately 75% of globally important crop types, including most fruits, seeds and nuts and several high-value commodity crops such as coffee, cocoa and oilseed rape^{1,7}. An estimated 5–8% of global crop production would be lost without pollination services, necessitating changes in human diets and a disproportionate expansion of agricultural land to fill this shortfall in crop production by volume⁸. Over the past 50 years, yields of crops with greater pollinator dependence have increased at a lower rate, and become more variable than crops that are less pollinator-dependent, suggesting that pollination services can be compromised by pollinator decline⁹. However, these estimates are often based on broad categorizations of pollinator dependence¹ derived from older, less standardized literature. A better understanding of the relationships between pollination services and crop productivity is therefore essential to quantify correctly how changing pollinator populations or diversity will affect food production.

Recent research indicates that pollinator-dependent crop productivity is important for balanced human diets. Pollinator-dependent crops are the principal sources of many micronutrients, including vitamins A and C, calcium, fluoride and folic acid¹⁰. Nutritional dependency on pollination overlaps geographically with the incidence of malnutrition of these nutrients. For example, areas with a high vitamin A deficiency are estimated to be three times more reliant on pollinator-dependent crops for plant-based vitamin A¹¹. Pollinator losses could therefore result in a substantial rise in the global rate of preventable diseases, such as ischaemic heart disease, potentially resulting in around 1.4 million additional deaths per year and approximately 29 million lost years of healthy life¹⁰. Although

¹Centre for Agri-Environmental Research, School of Agriculture, Policy and Development, University of Reading, Reading RG6 6AR, UK. ²Ecology Department, Biosciences Institute, S. Paulo University, 05508-090 S. Paulo Brazil and Vale Institute of Technology Sustainable Development, Belém 66055-090, Brazil. ³Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), IPBES Secretariat, UN Campus, Platz der Vereinten Nationen 1, Bonn D-53113, Germany. ⁴Laboratorio Ecotono, INIBIOMA-CONICET and Centro Regional Bariloche, Universidad Nacional del Comahue, Quintral 1250, 8400 San Carlos de Bariloche, Rio Negro, Argentina. ⁵Naturalis Biodiversity Center, PO Box 9517, Leiden 2300 RA, The Netherlands. ⁶Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam P.O. Box 94248, The Netherlands. ⁷School of Biological Sciences, University of East Anglia, Norwich NR4 7TL, UK. ⁸Instituto de Investigaciones en Recursos Naturales, Agroecología y Desarrollo Rural (IRNAD), Sede Andina, Universidad Nacional de Río Negro (UNRN), and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Mitre 630, CP 8400, San Carlos de Bariloche, Río Negro, Argentina. ⁹CSIRO Land and Water and James Cook University Division of Tropical Environments and Societies, Box 12139, Earlville BC, Cairns, Queensland 4870, Australia. ¹⁰Helmholtz Centre for Environmental Research—UFZ, Department of Community Ecology, Theodor-Lieser-Strasse 4, 06210 Halle, Germany. ¹¹Div. German Centre for Integrative Biodiversity Research, Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany. ¹²NERC Centre for Ecology and Hydrology, Bush Estate, Penicuik, Edinburgh EH26 0QB, UK.

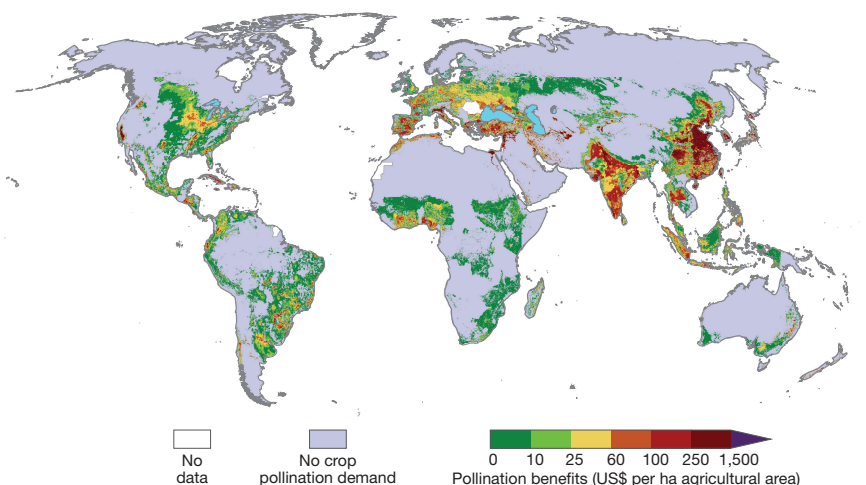


Figure 1 | Pollination service contribution to the crop market output in terms of US\$ per hectare of added production. Benefits are given for the year 2000 and have been corrected for inflation (to the year 2009) and for purchasing power parities. Plotted on a 5° by 5° latitude–longitude grid. Figure is adapted from ref. 4.

these studies provide estimates of qualitative risks, accounting for the distribution of food throughout the world would result in a better understanding of the effect of pollinator declines on human diets and health.

Animal pollination substantially affects global crop markets, with animal-pollinated crops often having higher sale prices than pollinator-independent crops. On the basis of 2009 market prices and production, animal pollination services enhance global crop output by an additional US\$235–577 billion (inflated to 2015 US\$) annually⁴. These economic benefits are unevenly distributed, with the greatest benefits in southern and eastern Asia and Mediterranean Europe, owing to greater production of highly pollinator-dependent crops and higher market prices (Fig. 1). The resultant price rises could reduce crop consumer welfare by an estimated US\$160–191 billion (inflated to 2015 US\$) and welfare in other, related markets such as agrochemicals, by a further US\$472–546 billion¹² (inflated to 2015 US\$) owing to shifts in crop production patterns towards lower input crops and non-agricultural activity.

The estimated relative economic impacts of pollinator losses on local producer and consumer welfare are greatest in western, northern and central Africa¹², highlighting a mismatch between total economic benefits and local effects on producers and consumers. Currently, economic analyses are limited by ecological and economic data gaps. Emerging methods, using holistic agro-ecological data, can redress this by directly linking marginal pollinator population shifts to crop output and welfare¹³, supporting more precise decision-making at local scales.

Many of the most widely grown and valuable cash crops, such as cocoa, almonds and coffee, are animal-pollinated¹; they provide employment and income for millions. Agriculture employs 1.4 billion people, approximately one-third of the world's economically active labour force¹⁴. This is particularly important to the world's poorest rural communities, 70% of whom rely on agriculture as the main source of income and employment¹⁵. More than 2 billion people in developing nations (83% of the global agricultural population) are reliant on smallholder agriculture, an area that has been largely neglected in pollinator research¹⁶. A recent study across small and large farms from Africa, Asia and Latin America found that, in fields that are 2 hectares or less, yield gaps owing to pollination deficits could be closed by a median of 24% through higher pollinator density. In larger fields, such benefits only occurred at high pollinator species richness¹⁷. This highlights the positive links between pollinators and yields in small- and large-holding crop systems worldwide.

Beyond food provisioning, pollinator-dependent plants contribute directly to medicines, biofuels, fibres, construction materials, musical instruments, arts, crafts and recreation activities. A case study in India found that 40% of plants that provide non-timber forest products, including medicine and construction materials, benefit to some extent from pollination

services¹⁸. The use of animal pollinated biofuel crops is growing, with the cultivation area of oilseed rape, sunflowers and soybeans increasing by 4.2 million hectares (32%) across Europe between 2005 and 2010 (ref. 19).

Bees can help to ensure livelihood security and alleviate poverty among rural communities through honey-hunting and beekeeping practices based on indigenous and local knowledge, documented in more than 50 countries^{20,21}. These practices typically require minimal investment, generate diverse saleable products, can occur often without land ownership or rent, provide flexibility in timing and locations of activities, link to culture and traditions, and produce family nutrition and medicinal benefits²². Anti-bacterial, anti-fungal, and anti-diabetic agents can be derived from honey; a recent review has found evidence that honey, as a topical treatment, can heal burns more quickly than conventional dressings²³.

Pollinators and their products also benefit society indirectly as sources of inspiration for art, music, literature, religion, traditions, technology and education. Bees inspire texts and imagery in many global religions, with examples including the *Surat An-Nahl* in the Qur'an, the three-bee motif of Pope Urban VIII and sacred passages within Hinduism, Buddhism and Chinese traditions such as the Chuang Tzu. For many people, a good quality of life arises from the role of pollinators as symbols of identity, from aesthetically important flowers in landscapes²⁴, in social relations, and globally significant heritage. Many sites listed under the *Convention Concerning the Protection of the World Cultural and Natural Heritage* depend on pollination to maintain their values, including the Classic Gardens of Suzhou in China and the *Agave Landscape and Ancient Industrial Facilities of Mexico*. The *Convention for the Safeguarding of the Intangible Cultural Heritage* recognizes several practices that rely on pollinator-dependent plants as internationally important, including the Argan practices and know-how concerning the argan tree (*Argania spinosa*) from Morocco, and Kimjang, making and sharing kimchi from the Republic of Korea.

At present, markets and economic indicators fail to capture all benefits from pollinators, and the full costs of supporting managed pollinators. As many land-use decisions rely on market forces and economic indicators, such failures can result in sub-optimal land management decisions that erode these benefits. Integrated monetary and non-monetary valuations of pollinator gains and losses can better inform decision-making on land-use, but will require the use of transdisciplinary methods, such as deliberative multi-criteria cost–benefit analysis that supports consideration of trade-offs among multiple dimensions.

Status and trends of pollinators and pollination

Information for assessing trends for wild pollinators comes from two main sources. First, historical information from museum collections and

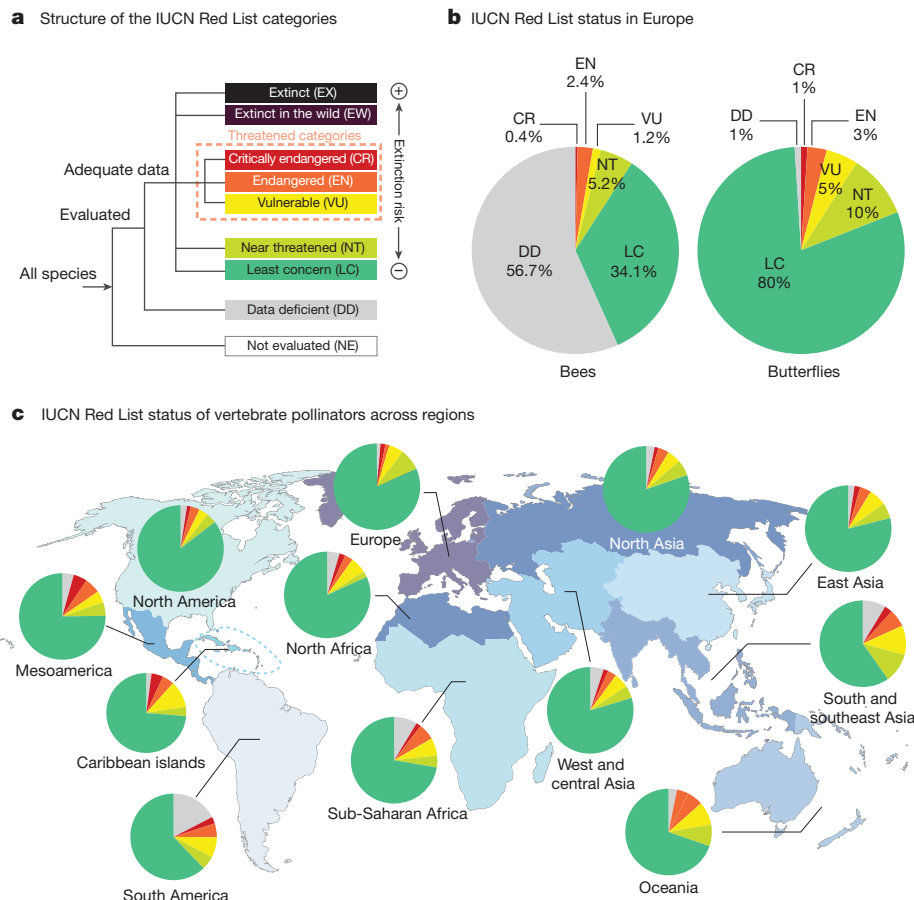


Figure 2 | The International Union for Conservation of Nature (IUCN) Red List status of wild pollinator taxa. a, Standardised IUCN extinction risk categories. **b,** European bees and butterflies. **c,** Vertebrate pollinators (including mammals and birds) across IUCN regions. IUCN relative risk

categories: EX, extinct; EW, extinct in the wild; CR, critically endangered; EN, endangered; VU, vulnerable; NT, near threatened; LC, least concern; DD, data deficient; NE, not evaluated. Figure adapted from ref. 5.

records collected by amateur naturalists and scientists; second, recently initiated surveys responding to concerns about pollinator declines. Although data are available only for some pollinator groups and for a few global regions, evidence is clear that several species have reduced their geographical ranges, a handful have gone extinct, and many have shown declines in local abundance. Global International Union for Conservation of Nature (IUCN) Red List assessments for vertebrate pollinator species (mostly birds and bats) estimate that 16.5% are threatened with global extinction (increasing to 30% for island species), trending towards more extinctions²⁵ (Fig. 2). For insect pollinators, the most important group of pollinators in the majority of biomes, a regional Red List assessment is available only for Europe, which indicates that 9% of bees²⁶ and 9% of butterflies²⁷ are threatened (Fig. 2). The figure for bees may increase substantially when more species are evaluated because the lack of data currently precludes assessment of 57% of European species. National Red Lists for bees are available for several European countries and indicate that up to 50% of bee species are nationally threatened²⁶.

Declines in bee diversity over the last century have been recorded in highly industrialized regions of the world, particularly northwestern Europe and eastern North America^{28–33}. Several bumblebee species have severely declined in occurrence, for instance Franklin's bumblebee (*Bombus franklini*) in the western United States, Cullum's bumblebee (*Bombus cullumanus*) in Europe and the giant bumblebee of Patagonia (*Bombus dahlbomii*)⁵. Declines in diversity in some areas, for example the Netherlands and the United Kingdom, seem to have slowed down in recent decades, but populations are still far below pre-decline levels of the early twentieth century³². Pollinators are also shifting ranges to more temperate latitudes or higher altitudes, following climate change, but often seem unable to track temperature shifts completely. Northern

Hemisphere bumblebees, for example, are losing suitable habitat at their southern range limits but are not expanding consistently at their northern range limits³⁴.

The most widespread managed pollinator is the western honeybee (*A. mellifera*) and globally the number of hives has increased by 45% during the last five decades, despite a temporary drop during the 1990s after the dissolution of the Soviet Bloc³⁵. National trends vary widely among countries (Fig. 3); for example there were recent declines in the United States and Germany but large increases in China, Argentina and Spain during the same period. More recently, many countries around the world report large-scale seasonal losses^{36,37}. National and local declines may not necessarily lead to more long-term colony losses as in some cases beekeepers can, at an economic cost, make up for seasonal colony loss by splitting colonies later in the season.

Shifts in wild and managed pollinator abundance, community diversity and composition may lead to shifts in flower visitation and ultimately affect fruit or seed set in wild plants and crops. The impact depends on the level of redundancy reflected in plant–pollinator interaction networks. Large and well-connected plant–pollinator networks are more likely to provide acceptable levels of pollination to plants as well as sufficient floral resource availability to pollinators. Networks show considerable changes in space and time, but whether and when networks suffer irreversible change following pollinator losses is not well understood (but see ref. 38).

Loss of both wild and managed pollinators may negatively affect human food production as many crop types rely, at least to some extent, on animal pollination for the quantity and/or quality of their yield¹. The production of crops that depend directly on pollinators constitutes a small proportion of the global food volume (5–8%). However, despite being small, the fraction of total agricultural production that depends directly on pollinators

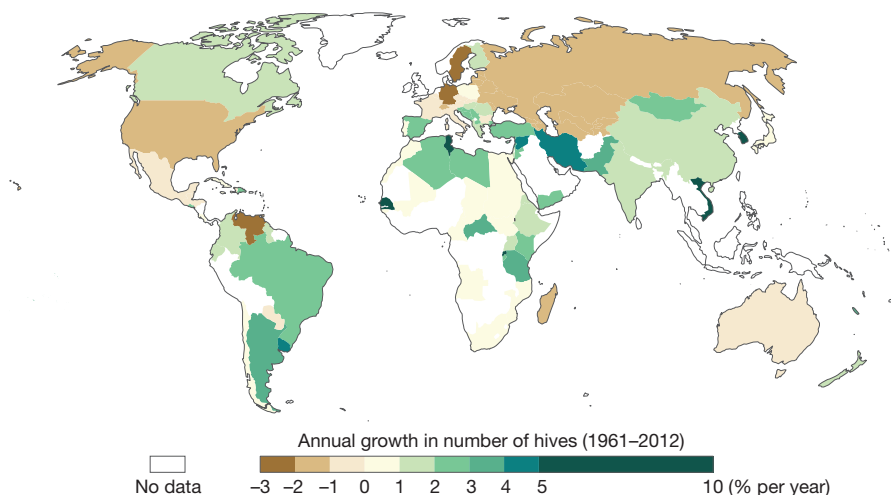


Figure 3 | Annual growth rate (percentage per year) in the number of honeybee hives for countries reporting data to the Food and Agriculture Organization (FAO) between 1961 and 2012. Data values are displayed at the country level; however, the distribution of honeybee hives is spatially heterogeneous within countries. Figure is based on data from the FAO (<http://faostat.fao.org/>, 2013).

has increased fourfold over the last five decades compared to a twofold increase in the fraction that does not depend on pollinators^{5,8} (Fig. 4). Consequently, global agriculture is now twice as dependent on pollinators compared to five decades ago. There are strong regional patterns in the pollinator-dependence of agriculture, with higher dependence in countries growing cash crops such as coffee, almonds, cocoa, soybeans or rapeseed

at large scales⁴. Crops that depend on pollinators have experienced the fastest global expansion in cultivated area and account for most of the approximately 30% expansion of global agricultural land during these five decades³⁹. Although yields show growth in most crops, owing to technological developments, pollinator-dependent crops have exhibited a slower average growth in yield, and higher inter-annual yield variability, than

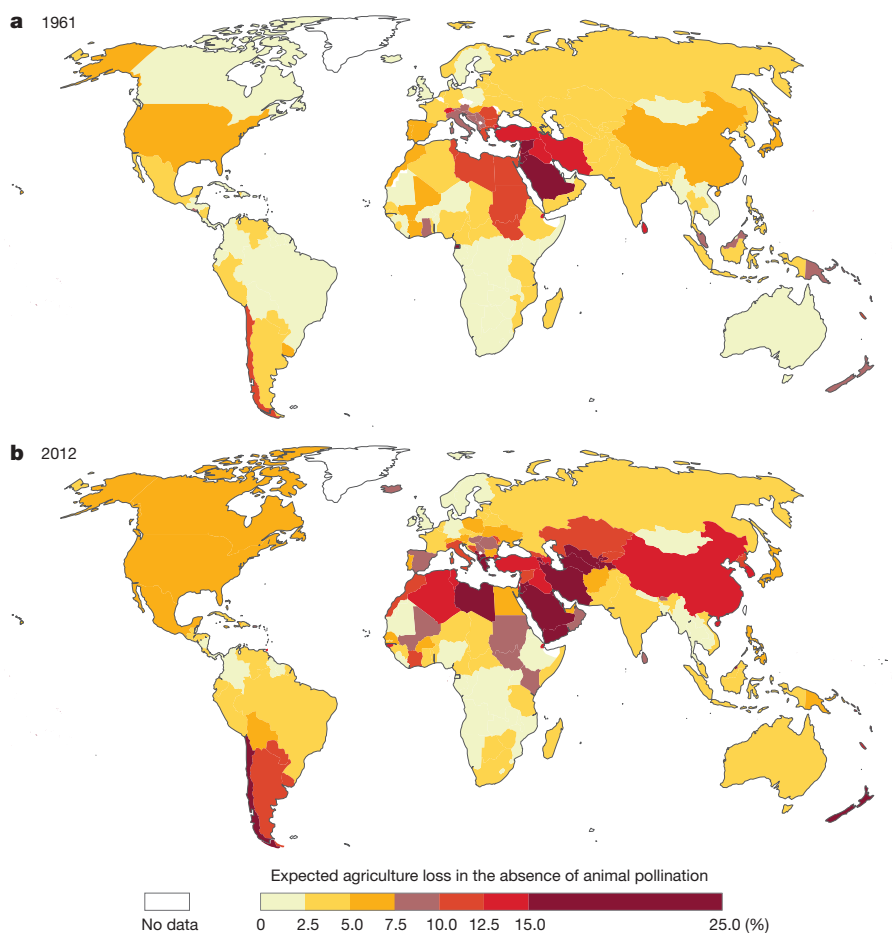


Figure 4 | Agriculture dependence on pollinators in 1961 and 2012. Figure is based on the FAO data set (<http://faostat.fao.org/>) and following the methodology of ref. 8. Data values are estimated at the country level; however the distribution of agricultural land, different crops and, thus, values of pollinator dependence are spatially heterogeneous within countries.

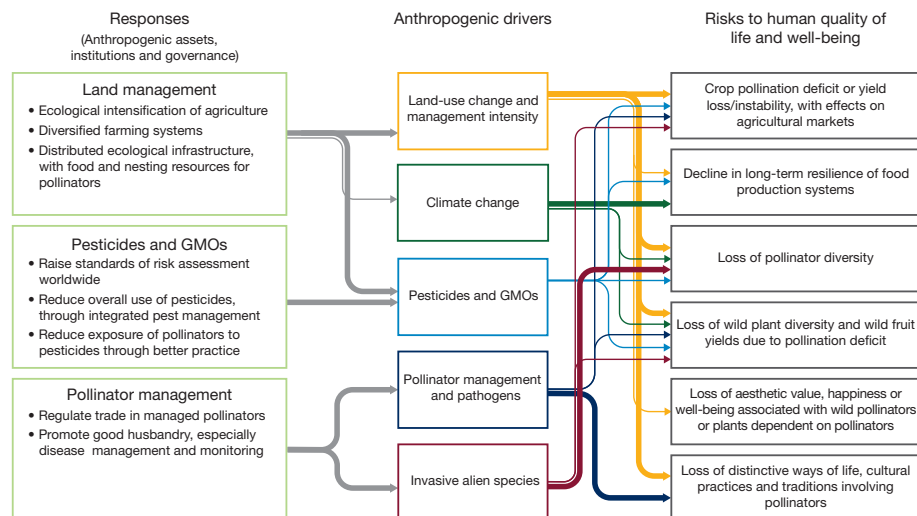


Figure 5 | Drivers, risks and responses to pollinator decline. Drivers of pollinator decline (central boxes) relate to the key risks associated with pollinator decline (right boxes), and how these drivers are addressed by three important sets of responses (left boxes) that ultimately reduce the risks. Responses combine elements of human facilities, knowledge, infrastructure and technology ('anthropogenic assets') with institutions and governance⁵. Arrows are thick if there is clear evidence that at least one of the responses can reduce the impact of the driver on pollinators,

or clear evidence that the driver generates the impact underlying the risk, at least in some circumstances. Arrows are thin if the link between response and driver, or driver and risk, is suspected or inferred by current evidence, but direct empirical evidence of it taking place is either sparse or lacking. This list of responses to pollinator decline is not exhaustive. There are 74 responses listed in ref. 5. Many responses also represent opportunities to improve livelihoods and environments directly. GMOs, genetically modified organisms.

pollinator-independent crops⁹. At the field level, decreased crop yield is related to lower abundance and diversity of pollinators in many crops^{3,17}. A higher density of flower visitors was the single most important factor improving yield in a study covering 33 pollinator-dependent crop systems across three continents¹⁷.

A loss of pollinators may have negative impacts on the reproduction of wild plants, as more than 90% of tropical flowering plant species and about 78% of temperate-zone species rely, at least in part, on animal pollination⁴⁰. There is a lack of data on large-scale and long-term trends in pollination or seed production, although historical shifts in plant distributions have been documented. Wild plants in the Netherlands and United Kingdom that require bees for cross-pollination showed declines corresponding to those of the bees that pollinate them²⁸. Similarly, declines are greatest for bee species that depend on forage plants that are also showing declines⁴¹. Although correlative, these patterns strongly suggest that plant and pollinator shifts are interdependent. Historical pollination rates can rarely be assessed, owing to the lack of quantitative data. However, analysis of pressed herbarium-specimens of pollinator-dependent orchids revealed that pollination had dropped from more than 40% to almost 0% in a century in Africa⁴². Further research is needed to understand the impacts of pollinator loss on wild plant communities and the other way around, particularly because a large part of wider biodiversity depends on the fruits, seeds and plant communities that pollinators maintain.

While shifts in pollinator diversity and ranges are relatively well-documented in parts of Europe and North America, a lack of wild pollinator data (species identity, distribution, occurrence and abundance) for Latin America, Africa and Asia limit any general conclusions on regional status and trends⁴³. Only long-term international and national monitoring of both pollinators and pollination can provide information for the majority of species and most parts of the world.

Drivers of change and responses

Comprehensibly linking observed long-term pollinator declines^{29,30,32} with specific or multiple drivers is not often possible owing to the lack of data on the status of pollinators (but see refs 44,45). Nonetheless, a wealth of studies worldwide point to a high likelihood that several anthropogenic drivers are threatening the abundance, diversity and health of wild and managed pollinators, and the pollination services they provide to wild

plants and crops^{46,47}. Figure 5 shows the five major drivers of pollinator decline identified in the literature.

Some drivers generate many stressors for pollinators (for example, increased land management intensity leads to loss of habitat and reduced floral resource supply), whereas interactions between drivers (for example, land-use, pesticides and climate change) may increase overall impact^{46,47}. The drivers create risks to human quality of life and well-being by eroding the benefits that humans obtain from pollinators and their pollination services (Fig. 5). From a scientific technical perspective, a risk is usually understood as the probability of a specific hazard or effect taking place, and risks are evaluated by estimating both the probability and the size or scale of the impact. Here we identify six risks to human well-being generated by pollinator decline. Given the substantial knowledge gaps regarding the status, trends and drivers of change in pollinators in most regions of the world, it is not yet possible to evaluate or rank these different risks quantitatively. Instead, we provide an overview of what is known about the link between each driver and risk (summarized in Fig. 5). For each major driver, we consider available responses and summarize what is known about their likely effectiveness (Fig. 5).

Changes in land-use and management intensity

Pollinators and pollination services are threatened by land-use changes involving the destruction, fragmentation and degradation of semi-natural habitats or the conversion of diversified farming systems into conventional intensive agriculture (that is, large, homogenous fields with high agrochemical inputs and intensive tillage, grazing or mowing)^{46,47}. Such land-use changes and management intensification can reduce or modify the supply of floral (pollen and nectar) and nesting resources to pollinators^{28,41,48}, often leading to lowered pollinator density or diversity and homogenized pollinator community structure^{45,49}. Changes to pollinator floral resources in contemporary intensive landscapes⁴⁸ could lead to malnutrition of individuals and colony stress, and increase vulnerability to pesticides and pathogens (see below) through impaired ability of the insect immune system to break down toxins in the diet^{47,50}.

Mitigating the pollinator-associated risks to humans driven by land-use change (Fig. 5) requires strategies that reverse or slow landscape homogenization. Current management options that enhance pollinator diversity or foraging densities at local and, to a lesser degree, landscape scales^{51–53}

BOX 1

Three complementary approaches to safeguard pollinators in agro-ecosystems

Ecological intensification⁶¹

This involves field and landscape management to increase the intensity of ecosystem services, such as biotic pest regulation, nutrient cycling and pollination, to enhance agricultural productivity and reduce reliance on agro-chemicals. Some specific actions to achieve ecological intensification are those that improve conditions for pollinators, such as creating flower-rich field margins⁵⁶. However, under ecological intensification, these actions are designed to facilitate on-farm pollination of particular crops. Promoting ecological intensification via agricultural extension services to demonstrate applications and convey the economic value of pollination services would probably increase adoption of beneficial practices⁸⁰. However, there are gaps in scientific understanding of the extent to which ecological intensification can assure farm yields (but see ref. 17), increase profitability at the farm scale, or which practices are the most effective to achieve these outcomes.

Strengthening diversified farming systems

Measures (tailored to the particular landscape) such as intercropping, using crop rotations that include flowering crops, agroforestry, managing forest or home gardens, and creating, restoring or maintaining native wild flower habitats, can be expected to foster diverse pollinator communities and pollination^{111,112}. Existing diversified farming systems are supported by many indigenous peoples and local communities worldwide, based on their cultural practices and knowledge systems. For example, central American *milpa* systems contain diverse plant communities attractive to insects⁶⁵ and home gardens in Mexico increase the fruit set of the columnar cactus *Stenocereus stellatus*¹¹³. Shifting cultivation systems under some circumstances support a diversity and abundance of floral resources for pollinators, through practices including integrating crops with flowering trees, customary rules that protect pollinator habitat¹¹⁴ and incorporation of wild plants in the production system¹¹⁵.

Ecological infrastructure

This infrastructure, needed to benefit crop pollination services, comprises small- to medium-sized patches of natural or semi-natural habitat (<10 ha), distributed throughout agricultural landscapes providing nesting and floral resources within reach of foraging pollinators (typically 500 m to 1 km between patches). The same approach should benefit pollinators and crop pollination in urban areas, although both agricultural and urban landscapes providing larger patches of natural habitat will probably increase regional pollinator diversity^{116,117}. Road verges, power lines, railway banks and waterways within and between urban areas have a large potential to become valuable infrastructure for supporting pollinators, if managed appropriately to provide flowering and nesting resources¹¹⁸. Connecting pollinator habitat patches together with such linear features enables movement of pollinators and can also enhance pollination of wild plants^{119,120}, although its role in maintaining pollinator populations remains unclear.

include organic farming practices^{49,54,55} and planting flower strips that provide floral resources⁵⁶. The efficacy of these measures, however, tends to be greatest in landscapes dominated by intensive agriculture, offering few floral resources^{54,56}. We argue that achieving sustainable, productive agriculture commensurate with pollinator and wider biodiversity on a large scale will require three complementary approaches, namely: (i) ecological intensification; (ii) strengthening diversified farming systems; and (iii) investing in ecological infrastructure by creating patches of (semi-)natural habitat throughout landscapes (Box 1).

There are potential trade-offs between maximising crop yield through conventional means and improving conditions for pollinator biodiversity with enhanced delivery of pollination services⁵⁷. For example, many farming systems using current organic practices usually produce lower crop yields⁵⁸, although some studies have demonstrated that yields of insect-pollinated crops can be improved through enhanced pollination services under organic management^{59,60}. Similarly, providing ecological infrastructure, such as flower strips or other habitat, may risk overall yield reductions by taking farmland out of direct crop production. To date, however, the only study to test this found that over several years the yield

lost at the farm scale was balanced by a yield gain in pollinator-dependent crops grown in rotation⁶¹.

Most of the scientific evidence on the efficacy of organic farming and growing flower strips in aiding pollinator biodiversity and pollination services is from Europe and North America^{55,56}, so caution is needed in extrapolating to other regions that differ in farm practices, underlying biodiversity and landscape character. Nonetheless, across the world managing the farm or landscape for wild pollinators coupled with use of managed bees results in the highest yields for animal-pollinated crops³. These improvements in pollination are partly due to higher local pollinator densities and partly due to higher pollinator species diversity providing complementarity or redundancy in service provision^{3,17}.

The policies and practices involved in ecological intensification, strengthening diversified farming systems and maintaining or providing ecological infrastructure can also have wider livelihood benefits for rural communities. These benefits can be fostered through holistic responses including food sovereignty⁶², and biocultural conservation approaches⁶³ that recognize rights, support economies and address negative multipliers of land-use change effects such as loss of access to traditional territories and loss of traditional knowledge⁶⁴. Many of the 32 Globally Important Agricultural Heritage Systems are based on indigenous and local knowledge that support the roles of pollinators in maintaining plant diversity, such as for lemons in the pergola-growing lemon gardens of southern Italy. Indigenous peoples and local communities view some of their practices as supporting an abundance and diversity of pollinators, for example favouring heterogeneity in landscapes and gardens, fostering pollinator resources, and recognizing kinship relationships that require respect and care for pollinators⁶⁵.

Pesticides

The risk to pollinators from pesticides (including insecticides, acaricides, fungicides, molluscicides and herbicides) is through a combination of toxicity and level of exposure. This risk varies according to the species' biology (for example, ability to metabolize toxins, foraging ecology), between chemical compounds, with the type and scale of land management^{66,67}, interactions with other stressors^{47,50}, and with landscape ecological infrastructure⁶⁸. Herbicides, used to control weeds, pose an indirect risk because they reduce the abundance and diversity of flowering plants providing pollen and nectar to pollinators⁶⁹.

Under controlled experimental conditions, pesticides, particularly insecticides like neonicotinoids, have a broad range of lethal and sublethal (for example, behavioural and physiological) effects on pollinators^{66,67,70,71} and, dependent on the concentration pollinators are exposed to, may reduce the pollination service that they provide⁷². The few available true field studies assessing the effects of field-realistic exposure on pollinators provide conflicting evidence of their effect based on the species studied and pesticide used⁶⁶. For example, a recent landscape-scale field experiment showed reduced wild pollinator survival and reproduction following actual field exposure to a neonicotinoid⁷³. However, evidence for negative effects on managed honeybees at the colony scale is often conflicting or its biological significance is contested^{66,67,70,71,73}.

Substantial gaps in knowledge remain regarding the effects of pesticides on pollinators. What constitutes actual field exposure is complicated by the ecology and foraging behaviours of different insect species in different land-use contexts^{66,67,71}. It is unclear whether sublethal effects on individual insects scale up to colonies and populations of managed bees and wild pollinators, especially over the longer-term^{66,70,71}. The potentially synergistic and long-term impacts of pesticide mixtures on insect individuals, colonies, and populations remain largely unresolved^{47,67,70,74}, although there is some correlational evidence linking reduced wild bee population persistence to neonicotinoid treatment of oilseed rape crops in England⁷⁵. Moreover, studies of sublethal insecticide effects have mainly tested a limited range of pesticides on a few bee species^{66,67,70,71}.

Despite the uncertainties, we can lower risks of pesticides to pollinators (and other non-target organisms) by decreasing levels of non-target toxicity and reducing exposure. Rigorous risk assessment of specific

pesticide ingredients and subsequent regulation have been shown to lower the overall environmental hazard from pesticides at a national scale^{76,77}. Regulatory risk assessments are usually only conducted on the western honeybee *A. mellifera*, which is not always a reliable surrogate for other pollinators, so broadening the set of species tested is one way of improving risk estimates⁴⁷. Pesticide exposure can be reduced by decreasing the usage of pesticides, for example, by adopting integrated pest management practices that only apply pesticide when pest pressure reaches an economic threshold⁷⁸. The effects of pesticides can be lessened through application practices, including the use of technologies to reduce pesticide spray drift⁷⁹ or decreasing herbicide use and adopting less stringent approaches to weed control, thereby allowing a variety of flowering weeds to flourish and support more diverse pollinator communities⁵⁴.

Education and training of farmers and the public are necessary to ensure the safe use of pesticides in agricultural and urban settings. Policy strategies that can help to reduce pesticide use, or avoid misuse, include supporting farmer field schools, known to increase the adoption of integrated pest management practices⁸⁰, and adopting national targets, codes of conduct or plans for risk reduction^{81,82}.

Research is needed (potentially co-developed with agri-business or farming communities) to provide viable alternatives to conventional high-chemical-input systems. The efficacy of pest management in pesticide-free and pesticide-minimized (for example, by integrated pest management) farming systems needs to be improved and the role of ecological intensification and/or ecological infrastructure (Box 1) in sustaining beneficial biodiversity while assuring farm profitability and yields needs to be studied. In all cases, protection of pollinators through reduced use or removal of very toxic chemicals from markets should be balanced against the need to ensure agricultural yields and food security.

Genetically modified crops

Most agricultural genetically modified organisms carry traits for herbicide tolerance or insect resistance. Although lethal risks from genetically modified crops to most pollinators are low, there are indirect risks from the management of these crops that need further evaluation. No lethal effects of insect-resistant crops from direct toxicity of pollen and nectar have been reported for bee species, although some have been observed in flower-visiting insects, such as beetles, butterflies and moths, that are closely related to the target pests⁸³. Management of herbicide-tolerant crops by eliminating weeds is likely to diminish floral resources for pollinators, although this remains little-studied⁸⁴. In one case, lower bee abundance in herbicide-tolerant fields led to an oilseed rape pollination deficit⁵⁹. Insect-resistant crops (for example, producing *Bacillus thuringiensis* toxins) can reduce insecticide use, although this varies with the crop and the prevalence of target and non-target pests in different regions. Reduced pesticide use accompanying insect-resistant crops could lower this pressure on non-target insects⁸⁵, but how it affects pollinators specifically is unknown. Moreover, the emergence of secondary outbreaks of non-target pests or primary pest resistance can lead to a resumption of pesticide use^{86,87}. The potential effects on pollinators of transgene flow and introgression (that is, gene substitution) in wild plant relatives and non-genetically modified crops require further study.

Environmental risk assessment is required for the commercial release of genetically modified crops in countries that are signatories of the Cartagena Protocol on Biosafety (Convention on Biological Diversity; <https://bch.cbd.int/protocol>). As for pesticides this regulatory framework uses a single pollinator species in toxicological assessment—the western honeybee *A. mellifera*. In addition, it does not address the sublethal or indirect effects of herbicide-tolerant and insect-resistant crops on pollinators⁸⁸. Further research would help inform the extent that policy responses are required in this area.

Pollinator management and pathogens

Insect pollinators suffer from a broad array of fungal, bacterial and viral pathogens and protozoan and invertebrate parasites^{46,47}. Host shifts mean that they represent a substantial current and future threat to the

health of bees, as seen with *Varroa* mites attacking and transmitting viruses among honeybees⁸⁹. The risk of disease and rates of pathogen and parasite transmission to new pollinator hosts has been exacerbated by the commercial management, mass breeding, and transport and trade in pollinators beyond their original ranges, which has increased the incidence of biological invasions^{89–91}. A notable example is the worldwide translocation of the western honeybee (*A. mellifera*) for its hive products and crop pollination services⁹⁰. This has resulted in a spill-over of pests and pathogens, both within *A. mellifera*—in the case of the *Varroa* mite originally picked up from *A. cerana*—and possibly between species, from *A. mellifera* to wild pollinators, such as the deformed-wing virus^{89,92}, although some pathogens may simply be shared generalists across flower visitors⁹³. It has also caused a decline in the cultural practices associated with keeping other bees native to particular areas, for example, stingless bees in central America⁹⁴. Commercial rearing of bumblebee species for crop pollination and their introduction to other continents has similarly resulted in biological invasions, pathogen transmission to native species and a decline of closely related congeners⁹¹.

Better regulation of national and international trade in managed pollinators species (and hive products), can help to limit the spread of parasites and pathogens and reduce the likelihood of ecological harm from further alien pollinator translocations. For instance, movement restrictions and biosecurity have so far prevented the establishment of *Varroa* mite into Australian populations of *A. mellifera*⁹⁵.

Although the epidemiology of pollinator communities needs to be studied, promoting good husbandry of managed bees will probably reduce pathogen spread across managed and wild bee populations⁹². Other options to minimize disease effects may include selective breeding for genetic diversity and resistance traits, whereas RNA interference technology may help treat *Varroa* and virus infestations⁹⁶.

Invasive alien species

The ecological effects of invasive alien species on pollinators and pollination are complex, but can be substantial under certain ecological and biogeographical circumstances. Invasive alien predators can transform ecosystems by consuming native pollinators, eliciting a shift to an invasive-dominated pollination system. For example, in Hawaii (United States) and the Ogasawara archipelago (Japan), the alien wasp (*Vespa pensylvanica*) and lizard (*Anolis carolinensis*) predators, respectively, drove native bees to extinction, leaving pollination reliant on alien honeybees^{97,98}. Alien plant or alien pollinator species modify native plant–pollinator networks, although the level of impact depends on the overlap in traits or niches, and at high abundances invasive alien pollinators can outcompete native pollinators^{91,98}. A noteworthy example is the decline and local extinction of the Patagonian giant bumblebee (*Bombus dahlbomii*) from much of its range following the introduction of the managed European bumblebee species (*B. terrestris* and *B. ruderatus*)⁹⁹, and related loss of raspberry production due to style damage from excessive flower visitation by *B. terrestris*¹⁰⁰. Invasive alien herbivores may indirectly disrupt pollination⁹⁸, and another potential risk to native plants comes from exotic plant pathogens, perhaps introduced with alien plants and spread by insects.

Eradication of established invasive aliens is seldom successful beyond oceanic islands, and often prohibitively expensive. Consequently, the most effective policy responses are surveillance and regulation to prevent new invasions, and rapid management once detected to avoid establishment.

Climate change

The effects of climate change on pollinators and pollination services may not be fully apparent for several decades¹⁰¹. However, over recent decades, the seasonal activity, abundance and range of some wild pollinator species (for example, bumblebees and butterflies) have been correlated with observed climate change^{102,103}. Most bumblebee distributions in Europe and North America are failing to track climate warming at the northern range limits of the species, leading to range contractions³⁴.

After 2050, all climate change scenarios reported by the Intergovernmental Panel on Climate Change suggest that the seasonal activity and ranges of many species may change differentially, disrupting life cycles and interactions between species and thus community composition¹⁰⁴. The rate of change of the climate across the landscape is expected to exceed the maximum speed at which many pollinator groups are able to colonize new areas^{34,101}. For some crops (such as apple and passion fruit) model projections at national scales have shown that these changes may disrupt crop pollination because of a future lack of overlap in the areas with the best climatic conditions for crops and their pollinators^{105–107}. Habitat loss and fragmentation may worsen this by limiting compensatory species migration¹⁰⁸, especially for species that are poor dispersers or habitat specialists^{28,44,109}. This may lead to increasingly species-poor plant–pollinator communities dominated by highly mobile habitat-generalist species⁴⁴.

Strategies to mitigate adverse effects of land-use change (Box 1 and Fig. 5), such as increasing crop diversity, regional farm diversity, and creating ecological infrastructure through targeted habitat conservation, creation or restoration, can help to secure pollinator diversity and pollination services for agriculture under climate change. Other measures, such as assisted translocation of pollinators to areas where pollination deficits arise, are untested and their effectiveness will probably be context-dependent¹¹⁰ and could pose threats to native pollinators.

Outlook

The potential risks to human well-being from pollinator decline mean that future research needs to understand which drivers and driver interactions are the most important in different contexts, to enable responses to be targeted. We encourage coordinated collaborative action and knowledge sharing to strengthen the establishment and implementation of the effective policy and management responses identified here to safeguard pollinators and pollination services for the long term.

Received 15 April; accepted 1 November 2016.

Published online 28 November 2016.

- Klein, A. M. *et al.* Importance of pollinators in changing landscapes for world crops. *Proc. R. Soc. B* **274**, 303–313 (2007).
- Discover Life's Bee Species Guide and World Checklist; http://www.discoverlife.org/mp/20q?guide=Apoidea_species&flags=HAS (Ascher and Pickering 2014).
- Garibaldi, L. A. *et al.* Wild pollinators enhance fruit set of crops regardless of honey bee abundance. *Science* **339**, 1608–1611 (2013).
This study is the first to show consistent benefits from wild insects to crop pollination across the globe and that those benefits cannot be replaced by increasing the abundance of a single managed species such as honeybees.
- Lautenbach, S., Seppelt, R., Liebscher, J. & Dormann, C. F. Spatial and temporal trends of global pollination benefit. *PLoS One* **7**, e35954 (2012).
This study is the most comprehensive and spatially explicit assessment of the direct economic benefits of pollination to global agriculture and accounts for differences in the effective spending power of different countries.
- Potts, S. G. *et al.* IPBES. *The Assessment Report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on Pollinators, Pollination and Food Production*; <http://www.ipbes.net/node/44781> (2016).
- Convention on Biological Diversity (CBD). *UNEP Decisions Adopted by the Conference of the Parties to the Convention on Biological Diversity at its Fifth Meeting (UNEP/CBD/COP/5/23/Annex III), Decision V/5*; <https://www.cbd.int/doc/decisions/COP-05-dec-en.pdf> (Nairobi, 2000).
- Gallai, N., Salles, J. M., Settele, J. & Vaissière, B. E. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecol. Econ.* **68**, 810–821 (2009).
- Aizen, M. A., Garibaldi, L. A., Cunningham, S. A. & Klein, A. M. How much does agriculture depend on pollinators? Lessons from long-term trends in crop production. *Ann. Bot.* **103**, 1579–1588 (2009).
This study is the first to take into account the partial dependence of most crops on pollinators in quantifying the effect of total loss of pollinators on global agricultural production, cultivated area and crop production diversity.
- Garibaldi, L. A., Aizen, M. A., Klein, A. M., Cunningham, S. A. & Harder, L. D. Global growth and stability of agricultural yield decrease with pollinator dependence. *Proc. Natl Acad. Sci. USA* **108**, 5909–5914 (2011).
- Smith, M. R., Singh, G. M., Mozaffarian, D. & Myers, S. S. Effects of decreases of animal pollinators on human nutrition and global health: a modelling analysis. *Lancet* **386**, 1964–1972 (2015).
- Chaplin-Kramer, R. *et al.* Global malnutrition overlaps with pollinator-dependent micronutrient production. *Proc. R. Soc. B* **281**, 20141799 (2014).
- Bauer, D. M. & Wing, I. S. The macroeconomic cost of catastrophic pollinator declines. *Ecol. Econ.* **126**, 1–13 (2016).
This study represents the most complete assessment of the consumer welfare impacts of pollinator losses both within and beyond crop markets.
- Hanley, N., Breeze, T. D., Ellis, C. & Goulson, D. Measuring the economic value of pollination services: principles, evidence and knowledge gaps. *Ecosyst. Serv.* **14**, 124–132 (2015).
- Altieri, M. A. Agroecology: the science of natural resource management for poor farmers in marginal environments. *Agric. Ecosyst. Environ.* **93**, 1–24 (2002).
- The World Bank. *Agriculture and rural development*; <http://data.worldbank.org/topic/agriculture-and-rural-development> (2015).
- Steward, P. R. *et al.* Pollination and biological control research: are we neglecting two billion smallholders. *Agric. Food Security* **3**, 5 (2014).
- Garibaldi, L. A. *et al.* Mutually beneficial pollinator diversity and crop yield outcomes in small and large farms. *Science* **351**, 388–391 (2016).
- Rehel, S. A. *et al.* Benefits of biotic pollination for non-timber forest products and cultivated plants. *Conserv. Soc.* **7**, 213–219 (2009).
- Breeze, T. D. *et al.* Agricultural policies exacerbate honeybee pollination service supply-demand mismatches across Europe. *PLoS One* **9**, e82996 (2014).
- Gupta, R. K., Reybroeck, W., van Veen, J. W. & Gupta, A. *Beekeeping for Poverty Alleviation and Livelihood Security. Vol. 1: Technological Aspects of Beekeeping* (Springer, 2014).
- Crane, E. *The World History of Beekeeping and Honey Hunting* (Routledge, 1999).
- Hilmi, M., Bradbear, N. & Mejia, D. *Beekeeping and sustainable livelihoods*. Food and Agriculture Organisation of the United Nations. Rural Infrastructure and Agro-Industries Division, Rome, Italy. (2011).
- Jull, A. B. *et al.* Honey as a topical treatment for wounds. *Cochrane Database Syst. Rev.* **3**, CD005083 (2015).
- Junge, X., Schüpbach, B., Walter, T., Schmid, B. & Lindemann-Matthies, P. Aesthetic quality of agricultural landscape elements in different seasonal stages in Switzerland. *Landsc. Urban Plan.* **133**, 67–77 (2015).
- Aslan, C. E., Zavaleta, E. S., Tershy, B. & Croll, D. Mutualism disruption threatens global plant biodiversity: a systematic review. *PLoS One* **8**, e66993 (2013).
- Nieto, A. *et al.* *European Red List of Bees* (Luxembourg: Publication Office of the European Union, Belgium, 2014).
- van Swaay, C. *et al.* *European Red List of Butterflies* (Luxembourg: Publication Office of the European Union, Spain, 2010).
- Biesmeijer, J. C. *et al.* Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science* **313**, 351–354 (2006).
- Cameron, S. A. *et al.* Patterns of widespread decline in North American bumble bees. *Proc. Natl Acad. Sci. USA* **108**, 662–667 (2011).
- Bommarco, R., Lundin, O., Smith, H. G. & Rundlöf, M. Drastic historic shifts in bumble-bee community composition in Sweden. *Proc. R. Soc. B* **279**, 309–315 (2012).
- Bartomeus, I. *et al.* Historical changes in northeastern US bee pollinators related to shared ecological traits. *Proc. Natl Acad. Sci. USA* **110**, 4656–4660 (2013).
- Carvalho, L. G. *et al.* Species richness declines and biotic homogenisation have slowed down for NW-European pollinators and plants. *Ecol. Lett.* **16**, 870–878 (2013).
This study uses 32 million data points to assess shifts in diversity of pollinator groups and plants in the Netherlands, the United Kingdom and Belgium over the last 80 years.
- Koh, I. *et al.* Modeling the status, trends, and impacts of wild bee abundance in the United States. *Proc. Natl Acad. Sci. USA* **113**, 140–145 (2016).
- Kerr, J. T. *et al.* Climate change impacts on bumblebees converge across continents. *Science* **349**, 177–180 (2015).
- Aizen, M. A. & Harder, L. D. The global stock of domesticated honey bees is growing slower than agricultural demand for pollination. *Curr. Biol.* **19**, 915–918 (2009).
- Neumann, P. & Carreck, N. L. Honey bee colony losses. *J. Apic. Res.* **49**, 1–6 (2010).
- Potts, S. G. *et al.* Declines of managed honeybees and beekeepers in Europe. *J. Apic. Res.* **49**, 15–22 (2010).
- Lever, J. J., van Nes, E. H., Scheffer, M. & Bascompte, J. The sudden collapse of pollinator communities. *Ecol. Lett.* **17**, 350–359 (2014).
- Aizen, M. A., Garibaldi, L. A., Cunningham, S. A. & Klein, A. M. Long-term global trends in crop yield and production reveal no current pollination shortage but increasing pollinator dependency. *Curr. Biol.* **18**, 1572–1575 (2008).
- Ollerton, J., Winfree, R. & Tarrant, T. How many flowering plants are pollinated by animals? *Oikos* **120**, 321–326 (2011).
- Scheper, J. *et al.* Museum specimens reveal loss of pollen host plants as key factor driving wild bee decline in the Netherlands. *Proc. Natl Acad. Sci. USA* **111**, 17552–17557 (2014).
- Pauw, A. & Hawkins, J. A. Reconstruction of historical pollination rates reveals linked declines of pollinators and plants. *Oikos* **120**, 344–349 (2011).
- Gill, R. J. *et al.* Protecting an ecosystem service: approaches to understanding and mitigating threats to wild insect pollinators. *Adv. Ecol. Res.* **54**, 135–206 (2016).
- Burke, L. A., Marlin, J. C. & Knight, T. M. Plant–pollinator interactions over 120 years: loss of species, co-occurrence, and function. *Science* **339**, 1611–1615 (2013).

45. Senapathi, D. *et al.* The impact of over 80 years of land cover changes on bee and wasp pollinator communities in England. *Proc. R. Soc. B* **282**, 20150294 (2015).
46. Potts, S. G. *et al.* Global pollinator declines: trends, impacts and drivers. *Trends Ecol. Evol.* **25**, 345–353 (2010).
47. Vanbergen, A. J. & the Insect Pollinators Initiative. Threats to an ecosystem service: pressures on pollinators. *Front. Ecol. Environ.* **11**, 251–259 (2013).
48. Baude, M. *et al.* Historical nectar assessment reveals the fall and rise of floral resources in Britain. *Nature* **530**, 85–88 (2016).
49. Kennedy, C. M. *et al.* A global quantitative synthesis of local and landscape effects on wild bee pollinators in agroecosystems. *Ecol. Lett.* **16**, 584–599 (2013).
- This study is based on detailed spatial modelling of data from 39 crop systems globally to understand the relative influence of landscape composition, landscape configuration, farm management and their interactions on wild bee abundance and richness.**
50. Collison, E., Hird, H., Cresswell, J. & Tyler, C. Interactive effects of pesticide exposure and pathogen infection on bee health – a critical analysis. *Biol. Rev. Camb. Philos. Soc.* **91**, 1006–1019 (2016).
51. Schneider, M. K. *et al.* Gains to species diversity in organically farmed fields are not propagated at the farm level. *Nat. Commun.* **5**, 4151 (2014).
52. Carvell, C., Bourke, A. F. G., Osborne, J. L. & Heard, M. S. Effects of an agri-environment scheme on bumblebee reproduction at local and landscape scales. *Basic Appl. Ecol.* **16**, 519–530 (2015).
53. Jönsson, A. M. *et al.* Sown flower strips in southern Sweden increase abundances of wild bees and hoverflies in the wider landscape. *Biol. Conserv.* **184**, 51–58 (2015).
54. Holzschuh, A., Steffan-Dewenter, I., Kleijn, D. & Tscharntke, T. Diversity of flower-visiting bees in cereal fields: effects of farming system, landscape composition and regional context. *J. Appl. Ecol.* **44**, 41–49 (2007).
55. Tuck, S. L. *et al.* Land-use intensity and the effects of organic farming on biodiversity: a hierarchical meta-analysis. *J. Appl. Ecol.* **51**, 746–755 (2014).
56. Scheper, J. *et al.* Environmental factors driving the effectiveness of European agri-environmental measures in mitigating pollinator loss—a meta-analysis. *Ecol. Lett.* **16**, 912–920 (2013).
- This is the first study to provide an overview of the effectiveness of a range of agri-environment options for supporting local pollinator richness and abundance.**
57. Blaauw, B. R. & Isaacs, R. Flower plantings increase wild bee abundance and the pollination services provided to a pollination-dependent crop. *J. Appl. Ecol.* **51**, 890–898 (2014).
58. Ponisio, L. C. *et al.* Diversification practices reduce organic to conventional yield gap. *Proc. R. Soc. B* **282**, 20141396 (2015).
59. Morandin, L. A. & Winston, M. L. Wild bee abundance and seed production in conventional, organic, and genetically modified canola. *Ecol. Appl.* **15**, 871–881 (2005).
60. Andersson, G. K. S., Rundlöf, M. & Smith, H. G. Organic farming improves pollination success in strawberries. *PLoS One* **7**, e31599 (2012).
61. Pywell, R. F. *et al.* Wildlife-friendly farming increases crop yield: evidence for ecological intensification. *Proc. R. Soc. B* **282**, 20151740 (2015).
- The first to test farm-scale ‘ecological intensification’, this study found no decrease in the crop yield over a five year rotation cycle on a large English farm, despite taking up to 8% of land out of production to support ecological functions.**
62. van der Ploeg, J. D. Peasant-driven agricultural growth and food sovereignty. *J. Peasant Stud.* **41**, 999–1030 (2014).
63. Gavin, M. C. *et al.* Defining biocultural approaches to conservation. *Trends Ecol. Evol.* **30**, 140–145 (2015).
64. Gómez-Baggethun, E. & Reyes-García, V. Reinterpreting change in traditional ecological knowledge. *Hum. Ecol.* **41**, 643–647 (2013).
65. Lyver, P., Perez, E., Carneiro da Cunha, M. & Roué, M. *Indigenous and Local Knowledge about Pollination and Pollinators Associated with Food Production: Outcomes from the Global Dialogue Workshop* <http://unesdoc.unesco.org/images/0023/002338/233811e.pdf> ((Panama 1–5 December 2014) UNESCO: Paris, 2015).
66. Godfray, H. C. J. *et al.* A restatement of the natural science evidence base concerning neonicotinoid insecticides and insect pollinators. *Proc. R. Soc. B* **281**, 20140558 (2014).
67. van der Sluijs, J. P. *et al.* Conclusions of the Worldwide Integrated Assessment on the risks of neonicotinoids and fipronil to biodiversity and ecosystem functioning. *Environ. Sci. Pollut. Res.* **22**, 148–154 (2015).
68. Park, M. G., Blitzer, E. J., Gibbs, J., Losey, J. E. & Danforth, B. N. Negative effects of pesticides on wild bee communities can be buffered by landscape context. *Proc. R. Soc. B* **282**, 20150299 (2015).
69. Gabriel, D. & Tscharntke, T. Insect pollinated plants benefit from organic farming. *Agric. Ecosyst. Environ.* **118**, 43–48 (2007).
70. Pisa, L. W. *et al.* Effects of neonicotinoids and fipronil on non-target invertebrates. *Environ. Sci. Pollut. Res.* **22**, 68–102 (2015).
71. Godfray, H. C. J. *et al.* A restatement of recent advances in the natural science evidence base concerning neonicotinoid insecticides and insect pollinators. *Proc. R. Soc. Lond. B* **282**, 20151821 (2015).
72. Stanley, D. A. *et al.* Neonicotinoid pesticide exposure impairs crop pollination services provided by bumblebees. *Nature* **528**, 548–550 (2015).
73. Rundlöf, M. *et al.* Seed coating with a neonicotinoid insecticide negatively affects wild bees. *Nature* **521**, 77–80 (2015).
- In a landscape experiment consisting of eight farms paired with controls, this study showed that actual field exposure to a neonicotinoid–pyrethroid seed treatment reduced wild bee densities, nesting success, bumblebee colony growth and reproduction, but did not measurably affect honeybee colony strength.**
74. Gill, R. J., Ramos-Rodriguez, O. & Raine, N. E. Combined pesticide exposure severely affects individual- and colony-level traits in bees. *Nature* **491**, 105–108 (2012).
75. Woodcock, B. A. *et al.* Impacts of neonicotinoid use on long-term population changes in wild bees in England. *Nat. Commun.* **7**, 12459 (2016).
76. Fischer, D. & Moriarty, T. *Pesticide Risk Assessment for Pollinators* (John Wiley & Sons, 2014).
77. Cross, P. Pesticide hazard trends in orchard fruit production in Great Britain from 1992 to 2008: a time-series analysis. *Pest Manag. Sci.* **69**, 768–774 (2013).
78. Ekström, G. & Ekblom, B. Pest control in agro-ecosystems: an ecological approach. *Crit. Rev. Plant Sci.* **30**, 74–94 (2011).
79. Johansen, E., Hooven, L. A. & Sagili, R. R. *How to Reduce Bee Poisoning from Pesticides* (Oregon State Univ. Extension Service, 2013).
80. Waddington, H. *et al.* Farmer field schools for improving farming practices and farmer outcomes in low- and middle-income countries: a systematic review. *Campbell Syst. Rev.* **10**, 1–335 (2014).
81. Barzman, M. & Dachbrodt-Saaydeh, S. Comparative analysis of pesticide action plans in five European countries. *Pest Manag. Sci.* **67**, 1481–1485 (2011).
82. Ekström, G. & Ekblom, B. Can the IOMC revive the ‘FAO code’ and take stakeholder initiatives to the developing world? *Outlooks Pest Manag.* **21**, 125–131 (2010).
83. Mommaerts, V., Jans, K. & Smagghe, G. Impact of *Bacillus thuringiensis* strains on survival, reproduction and foraging behaviour in bumblebees (*Bombus terrestris*). *Pest Manag. Sci.* **66**, 520–525 (2010).
84. Bohan, D. A. *et al.* Effects on weed and invertebrate abundance and diversity of herbicide management in genetically modified herbicide-tolerant winter-sown oilseed rape. *Proc. R. Soc. B* **272**, 463–474 (2005).
85. Marvier, M., McCreedy, C., Regetz, J. & Kareiva, P. A meta-analysis of effects of Bt cotton and maize on nontarget invertebrates. *Science* **316**, 1475–1477 (2007).
86. Lu, Y. *et al.* Mirid bug outbreaks in multiple crops correlated with wide-scale adoption of Bt cotton in China. *Science* **328**, 1151–1154 (2010).
87. Barfoot, P. & Brookes, G. Key global environmental impacts of genetically modified (GM) crop use 1996–2012. *GM Crops Food* **5**, 149–160 (2014).
88. Andow, D. A. *et al.* An ecologically-based method for selecting ecological indicators for assessing risks to biological diversity from genetically-engineered plants. *J. Biosaf.* **22**, 141–156 (2013).
89. Wilfert, L. *et al.* Deformed wing virus is a recent global epidemic in honeybees driven by *Varroa* mites. *Science* **351**, 594–597 (2016).
90. Moritz, R. F. A., Härtel, S. & Neumann, P. Global invasions of the western honeybee (*Apis mellifera*) and the consequences for biodiversity. *Ecoscience* **12**, 289–301 (2005).
91. Dohzono, I. & Yokoyama, J. Impacts of alien bees on native plant–pollinator relationships: a review with special emphasis on plant reproduction. *Appl. Entomol. Zool. (Jpn.)* **45**, 37–47 (2010).
92. Fürst, M. A., McMahon, D. P., Osborne, J. L., Paxton, R. J. & Brown, M. J. F. Disease associations between honeybees and bumblebees as a threat to wild pollinators. *Nature* **506**, 364–366 (2014).
- This paper sampled *Apis* and *Bombus* across 26 geographically dispersed sites in the United Kingdom revealing the co-prevalence of deformed wing virus (DWV) infections and evidence of local transmission, suggesting disease spill-over from managed honeybees to wild bumblebee species.**
93. Tehel, A., Brown, M. J. F. & Paxton, R. J. Impact of managed honey bee viruses on wild bees. *Curr. Opin. Virol.* **19**, 16–22 (2016).
94. Quezada-Euán, J. J. G., de Jesús May-Itzá, W. & González-Acereto, J. A. Meliponiculture in México: problems and perspective for development. *Bee World* **82**, 160–167 (2001).
95. Cook, D. C., Thomas, M. B., Cunningham, S. A., Anderson, D. L. & De Barro, P. J. Predicting the economic impact of an invasive species on an ecosystem service. *Ecol. Appl.* **17**, 1832–1840 (2007).
96. Hunter, W. *et al.* Large-scale field application of RNAi technology reducing Israeli acute paralysis virus disease in honey bees (*Apis mellifera*, Hymenoptera: Apidae). *PLoS Pathogens* **6**, e1001160 (2010).
97. Hanna, C., Foote, D. & Kremen, C. Invasive species management restores a plant–pollinator mutualism in Hawaii. *J. Appl. Ecol.* **50**, 147–155 (2013).
98. Traveset, A. & Richardson, D. M. Mutualistic interactions and biological invasions. *Annu. Rev. Ecol. Syst.* **45**, 89–113 (2014).
99. Morales, C. L., Arbetman, M. P., Cameron, S. A. & Aizen, M. A. Rapid ecological replacement of a native bumble bee by invasive species. *Front. Ecol. Environ.* **11**, 529–534 (2013).
100. Sáez, A., Morales, C. L., Ramos, L. Y. & Aizen, M. A. Extremely frequent bee visits increase pollen deposition but reduce drupelet set in raspberry. *J. Appl. Ecol.* **51**, 1603–1612 (2014).
101. Settele, J. *et al.* in *Climate Change 2014: Impacts, Adaptation, and Vulnerability* (eds Field, C. B. *et al.*) 271–359 (Cambridge Univ. Press, 2014).
102. Hegland, S. J., Nielsen, A., Lázaro, A., Bjerknes, A. L. & Totland, Ø. How does climate warming affect plant–pollinator interactions? *Ecol. Lett.* **12**, 184–195 (2009).

103. Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B. & Thomas, C. D. Rapid range shifts of species associated with high levels of climate warming. *Science* **333**, 1024–1026 (2011).
104. Schweiger, O., Settele, J., Kudrna, O., Klotz, S. & Kühn, I. Climate change can cause spatial mismatch of trophically interacting species. *Ecology* **89**, 3472–3479 (2008).
105. Giannini, T. C. *et al.* Identifying the areas to preserve passion fruit pollination service in Brazilian Tropical Savannas under climate change. *Agric. Ecosyst. Environ.* **171**, 39–46 (2013).
106. Polce, C. *et al.* Climate-driven spatial mismatches between British orchards and their pollinators: increased risks of pollination deficits. *Glob. Change Biol.* **20**, 2815–2828 (2014).
107. Settele, J., Bishop, J. & Potts, S. G. Climate change impacts on pollination. *Nat. Plants* **2**, 16092 (2016).
108. Forister, M. L. *et al.* Compounded effects of climate change and habitat alteration shift patterns of butterfly diversity. *Proc. Natl Acad. Sci. USA* **107**, 2088–2092 (2010).
109. Warren, M. S. *et al.* Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature* **414**, 65–69 (2001).
110. Burton, I. & Lim, B. Achieving adequate adaptation in agriculture. *Clim. Change* **70**, 191–200 (2005).
111. Frimpong, E. A., Gemmill-Herren, B., Gordon, I. & Kwabong, P. K. Dynamics of insect pollinators as influenced by cocoa production systems in Ghana. *J. Pollinat. Ecol.* **5**, 74–80 (2011).
112. Kremen, C. & Miles, A. Ecosystem services in biologically diversified versus conventional farming systems: benefits, externalities, and trade-offs. *Ecol. Soc.* **17**, 40 (2012).
113. Arias-Cóyotl, E., Stoner, K. E. & Casas, A. Effectiveness of bats as pollinators of *Stenocereus stellatus* (Cactaceae) in wild, managed *in situ*, and cultivated populations in La Mixteca Baja, central Mexico. *Am. J. Bot.* **93**, 1675–1683 (2006).
114. Padoch, C. & Pinedo-Vasquez, M. Saving slash-and-burn to save biodiversity. *Biotropica* **42**, 550–552 (2010).
115. Zizumbo-Villarreal, D., Vargas-Ponce, O., Rosales-Adame, J. J. & Colunga-GarcíaMarín, P. Sustainability of the traditional management of Agave genetic resources in the elaboration of mezcal and tequila spirits in western Mexico. *Genet. Resour. Crop Evol.* **60**, 33–47 (2013).
116. Hernandez, J. L., Frankie, G. W. & Thorp, R. W. Ecology of urban bees: a review of current knowledge and directions for future study. *Cities Environ.* **2**, 3 (2009).
117. Lopes, A. V., Girão, L. C., Santos, B. A., Peres, C. A. & Tabarelli, M. Long-term erosion of tree reproductive trait diversity in edge-dominated Atlantic forest fragments. *Biol. Conserv.* **142**, 1154–1165 (2009).
118. Berg, Å., Åhrné, K., Öckinger, E., Svensson, R. & Wissman, J. Butterflies in semi-natural pastures and power-line corridors – effects of flower richness, management, and structural vegetation characteristics. *Insect Conserv. Divers.* **6**, 639–657 (2013).
119. Van Geert, A., Van Rossum, F. & Triest, L. Do linear landscape elements in farmland act as biological corridors for pollen dispersal? *J. Ecol.* **98**, 178–187 (2010).
120. Kormann, U. *et al.* Corridors restore animal-mediated pollination in fragmented tropical forest landscapes. *Proc. R. Soc. B* **283**, 20152347 (2016).

Acknowledgements We thank Y. Estrada for preparing the figures and the authors of the original articles for providing the data that underpin them. We are grateful to the authors and reviewers of ref. 5 for their contributions to the report.

Author Contributions All authors contributed equally to the planning, evaluation of the literature and writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.G.P. (s.g.potts@reading.ac.uk).

Reviewer Information *Nature* thanks R. Gill, R. Paxton and N. Raine for their contribution to the peer review of this work.

Gamma frequency entrainment attenuates amyloid load and modifies microglia

Hannah F. Iaccarino^{1,3*}, Annabelle C. Singer^{2,3,4*}, Anthony J. Martorell^{1,3}, Andrii Rudenko^{1,3}, Fan Gao^{1,3}, Tyler Z. Gillingham^{1,3}, Hansruedi Mathys^{1,3}, Jinsoo Seo^{1,3}, Oleg Kritskiy^{1,3}, Fatema Abdurrob^{1,3}, Chinnakkaruppan Adaikkan^{1,3}, Rebecca G. Canter^{1,3}, Richard Rueda^{1,3}, Emery N. Brown^{1,3,5,6}, Edward S. Boyden^{2,3,4} & Li-Huei Tsai^{1,3,7}

Changes in gamma oscillations (20–50 Hz) have been observed in several neurological disorders. However, the relationship between gamma oscillations and cellular pathologies is unclear. Here we show reduced, behaviourally driven gamma oscillations before the onset of plaque formation or cognitive decline in a mouse model of Alzheimer's disease. Optogenetically driving fast-spiking parvalbumin-positive (FS-PV) interneurons at gamma (40 Hz), but not other frequencies, reduces levels of amyloid- β (A β)_{1–40} and A β _{1–42} isoforms. Gene expression profiling revealed induction of genes associated with morphological transformation of microglia, and histological analysis confirmed increased microglia co-localization with A β . Subsequently, we designed a non-invasive 40 Hz light-flickering regime that reduced A β _{1–40} and A β _{1–42} levels in the visual cortex of pre-depositing mice and mitigated plaque load in aged, depositing mice. Our findings uncover a previously unappreciated function of gamma rhythms in recruiting both neuronal and glial responses to attenuate Alzheimer's-disease-associated pathology.

Activation of local circuits of excitatory and fast-spiking inhibitory neurons that resonate at 20–50 Hz gives rise to oscillations in the local field potential (LFP), called gamma oscillations^{1–3}. Although studies have demonstrated disrupted gamma in various neurological diseases, the interplay between pathology and this emergent circuit property has yet to be determined^{4,5}. In general, molecular and cellular pathology is thought to alter synaptic activity. However, in at least one disorder, Alzheimer's disease (AD), changes in synaptic activity can also feedback to alter molecular pathology. Studies have shown that increases in synaptic activity *in vivo* increase levels of A β ⁶, a 36- to 43-amino-acid protein, whose aggregation is thought to initiate neurotoxic events, including neuroinflammation, synaptic and neuronal loss, and tau-associated pathology⁷. We aimed to determine how gamma affects molecular pathology in a mouse model of AD. Understanding how gamma might affect disease pathogenesis has important implications for elucidating both the basic pathology of and possible therapeutic interventions for neurological diseases with altered gamma.

Reduced hippocampal gamma in 5XFAD mice

Altered gamma has been observed in multiple brain regions in several neurological and psychiatric disorders, including a reduction in spontaneous gamma synchronization in patients with AD and reduced gamma power in multiple AD mouse models^{4,5,8,9}. However, it is unclear whether gamma is altered early in disease progression and whether it affects disease pathology. Accordingly, we recorded neural activity from behaving 5XFAD mice, a well-established model of AD¹⁰. In 3-month-old mice, which have elevated levels of A β but no major plaque accumulation in the hippocampus or manifestation of learning and memory deficits¹⁰, we recorded neural activity from hippocampal subregion CA1, where gamma has been particularly well characterized

(see, for example, refs 11–14), using a virtual environment (Extended Data Fig. 1a). In CA1, gamma is present during distinct periods of activity: running, when theta oscillations (4–12 Hz) occur (Extended Data Fig. 1b, left); and quiescent behaviour, when sharp-wave ripples (SWRs) occur^{15,16} (Extended Data Fig. 1b, right). We found no clear differences in slow gamma power (20–50 Hz) between 5XFAD mice and wild-type (WT) littermates during theta (Extended Data Fig. 1c, d).

We next examined gamma during SWRs, high-frequency oscillations of 150–250 Hz lasting around 50–100 ms (Extended Data Fig. 1b, e)¹⁴. Previous work has shown that slow gamma is elevated during SWRs and that increased gamma synchrony across CA3 and CA1 during SWRs correlates with more coordinated firing between neurons¹⁶. Similarly, we found increased gamma power during SWRs (Fig. 1a, a yellow arrow indicates elevated gamma, and Extended Data Fig. 1e). The instantaneous frequencies of these slower oscillations (10–50 Hz, Methods) were a unimodal distribution centred around 40 Hz (Fig. 1b and Extended Data Fig. 1f). Comparing gamma during SWRs in WT and 5XFAD littermates, we found that gamma power was significantly lower in 5XFAD than in WT mice (Methods, Fig. 1c and Extended Data Fig. 1g, j; examples in Fig. 1c, top). Spiking was phase-modulated by gamma in both groups, although the depth of modulation was significantly smaller in 5XFAD than in WT animals (Fig. 1d and Extended Data Fig. 1h, k). Furthermore, there were fewer SWRs per unit time in non-theta periods in 5XFAD mice compared with WT (Extended Data Fig. 1i), reducing periods when gamma power is elevated (Fig. 1a, c and Extended Data Fig. 1e). These results reveal deficits in gamma modulation of CA1 spiking in a mouse model of AD before the development of major amyloid plaque accumulation or evidence of cognitive impairment. This deficit in gamma converges with evidence of gamma deficits in different mouse models of AD, and

¹Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁴MIT Media Lab, Departments of Biological Engineering and Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁵Institute of Medical Engineering and Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁶Massachusetts General Hospital, Boston, Massachusetts, Massachusetts 02114, USA. ⁷Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02139, USA. Present addresses: Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Emory University, Atlanta, USA, 30332 (A.C.S.); Department of Biology, The City College of New York, New York, USA, 10031 (A.R.).

*These authors contributed equally to this work.

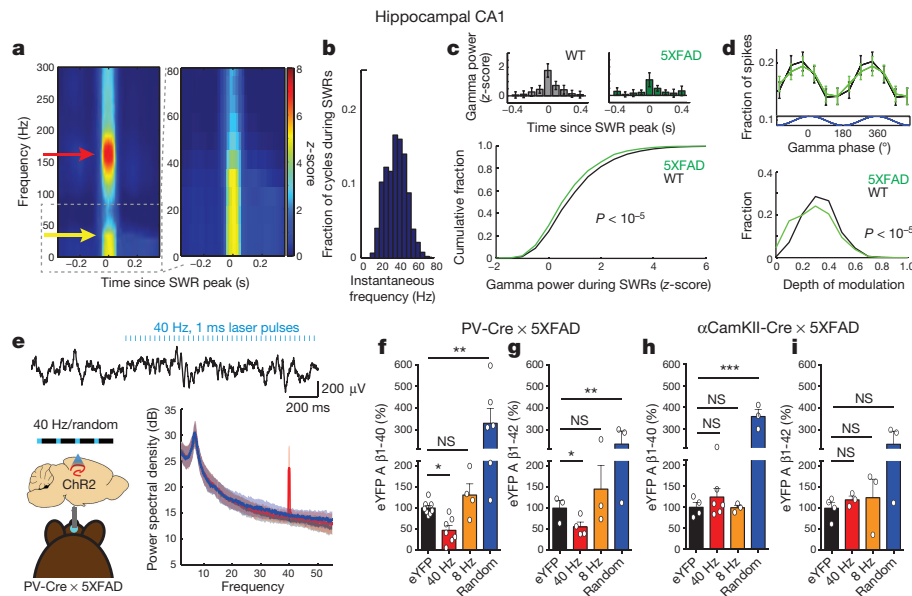


Figure 1 | 5XFAD mice have reduced power in gamma during hippocampal SWRs. **a**, Average SWRs-triggered spectrograms for one mouse (left) showing gamma (yellow arrow) during SWRs (red arrow); right: frequencies below 80 Hz enlarged ($n = 370$ SWR). **b**, Histogram of instantaneous gamma frequencies during SWRs for mouse in **a**. **c**, Above: z-scored gamma power around SWR peaks for one WT and one 5XFAD mouse (mean \pm s.e.m.). Below: cumulative distribution of gamma power during SWRs (rank-sum test, $n = 2,166$ and $3,085$ SWRs in six 5XFAD and WT mice, respectively). **d**, Above: fraction of spikes during SWRs as a function of gamma phase (mean \pm s.e.m.). Below: depth of gamma spiking modulation during SWRs (rank-sum test, bootstrap resampling, $n = 2,500$ 5XFAD and $3,000$ WT phase distributions). **e**, Above: LFP trace before

and during 40 Hz optogenetic stimulation. Below: mean and standard deviation of power spectral density ($n = 4$ 5XFAD and 3 WT mice). Red, 40 Hz; blue, random; black, no stimulation. **f**, Relative $A\beta_{1-40}$ levels in CA1 of 5XFAD/PV-Cre mice in each stimulation condition normalized to eYFP controls ($n = 8$ eYFP, $n = 7$ 40 Hz, $n = 4$ 8 Hz, $n = 6$ random mice). **g**, As in **f** for $A\beta_{1-42}$ ($n = 4$ eYFP, $n = 4$ 40 Hz, $n = 3$ 8 Hz, $n = 3$ random mice). **h**, Relative $A\beta_{1-40}$ levels in CA1 of 5XFAD/ α CamKII-Cre mice in each stimulation condition normalized to eYFP controls ($n = 6$ 40 Hz, $n = 3$ 8 Hz, $n = 3$ random mice). **i**, As in **h** for $A\beta_{1-42}$ ($n = 3$ mice per group). NS, not significant, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ by one-way analysis of variance (ANOVA); circles indicate n , mean \pm s.e.m. unless otherwise specified in bar graphs.

reports that gamma is altered in humans with AD^{5,8,9}. Indeed, molecular deficits in Nav1.1 in humans have been linked with gamma deficits in hAPP mice⁷.

Gamma reduced $A\beta$ production in CA1

These gamma deficits during SWRs early in disease progression in this mouse model of AD prompt the question of whether gamma could affect molecular and cellular AD pathophysiology. To test this, we induced gamma optogenetically in 5XFAD/PV-Cre mice (Methods, Fig. 1e, left, and Extended Data Fig. 2a–c). We chose to drive FS-PV-interneurons at 40 Hz because we found deficits in gamma during SWRs, and instantaneous gamma frequencies during SWRs were centred around 40 Hz. Delivering 1 ms, 473 nm light pulses at 40 Hz resulted in increased power at 40 Hz in LFPs in CA1, while random stimulation did not (Fig. 1e and Extended Data Fig. 1l). Both resulted in similar firing rates (Extended Data Fig. 1m–o).

$A\beta$ accumulation is thought to initiate multiple neurotoxic events typical for AD pathology. Therefore, we examined whether gamma stimulation affected overall $A\beta$ peptide levels in the hippocampus of 5XFAD mice. We found that 1 h of FS-PV-interneuron stimulation reduced $A\beta_{1-40}$ by 53.22% and $A\beta_{1-42}$ by 44.62% in the 40 Hz group, which expresses Chr2 compared with the enhanced yellow fluorescent protein (eYFP) control group, as measured in CA1 by $A\beta$ enzyme-linked immunosorbent assay (ELISA) (Fig. 1f, g and raw concentration (picograms per millilitre) in Extended Data Table 1). We performed a comprehensive set of control experiments to determine whether the effect was specific to frequency, cell type, and/or rhythmicity. Neither stimulation of CamKII-positive excitatory neurons at 40 Hz nor FS-PV-interneurons at 8 Hz or random intervals significantly reduced $A\beta$ levels (Fig. 1f–i and Methods). Immunohistochemical analysis using two β -amyloid-specific antibodies (Cell Signalling Technology; D54D2, BioLegend; 12F4 (ref. 17)) in CA1 confirmed these results: $A\beta$ labelling

intensity was significantly reduced by 39.5% after 40 Hz stimulation compared with eYFP controls (Fig. 2e, f, D54D2 antibody; Extended Data Fig. 2e, f, 12F4 antibody).

Brain amyloid concentration depends on $A\beta$ production from amyloid precursor protein (APP) and $A\beta$ clearance rates. To elucidate whether 40 Hz stimulation reduced $A\beta$ production, we examined its effects on APP cleavage by measuring levels of the cleavage intermediates of APP, C-terminal fragments (CTFs) and amino (N)-terminal fragments (NTF), in the hippocampus of the 5XFAD/PV-Cre mice. After 40 Hz stimulation, we found significantly reduced APP CTFs and NTF compared with eYFP and random controls (Fig. 2a–d and Extended Data Fig. 2d).

Previous work has shown that APP is transported and processed in recycling endosomes¹⁸, and enlarged early endosomes have been observed in brain tissue from patients with AD¹⁹. Therefore, we characterized endosomes in CA1 after stimulation using two markers, EEA1 (early endosomal antigen 1) and Rab5 (Ras-related protein encoded by the *RAB5A* gene). Altogether, the intensity of endosomal labelling of CA1 neurons significantly decreased in both EEA1 (39.7%) and Rab5 (40.1%) after 40 Hz stimulation compared with eYFP controls (Fig. 2e, g and Extended Data Fig. 2g, h). These results suggest that, in addition to observed changes in APP cleavage products, 40 Hz stimulation also alters general endosomal processing.

Gamma induced microglia morphological changes

To explore further the cellular and molecular effects of stimulation in an unbiased manner, we performed genome-wide RNA sequencing (RNA-seq) of CA1 tissue after 1 h of 40 Hz or no stimulation (eYFP) of the 5XFAD/PV-Cre mice (Fig. 3a and Extended Data Fig. 3a, b, c). Notably, 35% of all upregulated genes had their highest expression in microglia (Fig. 3b). This RNA-seq analysis strongly suggests that 40 Hz stimulation causes an alteration in the state of microglia, which

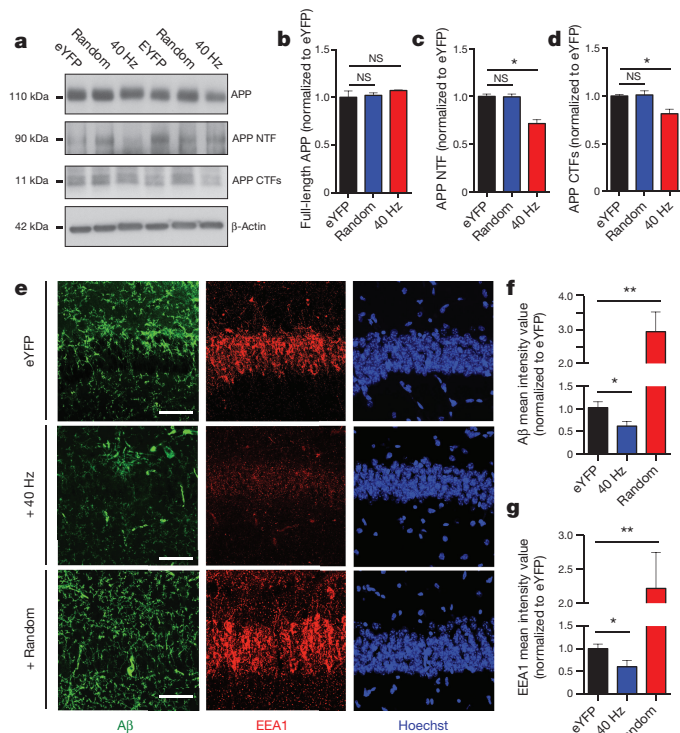


Figure 2 | Driving 40 Hz oscillations optogenetically in hippocampus reduces A β in 5XFAD mice. **a**, Representative western blot showing levels of APP (CT695), APP NTF (A8967), APP CTFs (CT695), and β -actin (A5316, loading control) in CA1 of 5XFAD/PV-Cre mice expressing only eYFP or Chr2 with 40 Hz, or random stimulation conditions. One mouse per lane, two biological replicates. **b**, Relative immunoreactivity of full-length APP normalized to actin (for **b–d**, $n = 6$ mice per group). **c**, Relative immunoreactivity of APP NTF normalized to actin. **d**, Relative immunoreactivity of APP CTFs normalized to actin. **e**, Immunohistochemistry with anti-A β (D54D2, green) and anti-EEA1 (610457, red) antibodies in CA1 of 5XFAD/PV-Cre mice (scale bar, 50 μ m). **f**, Relative immunoreactivity of A β normalized to eYFP controls (for **f**, $n = 3$ mice per group). **g**, Relative immunoreactivity of EEA1 normalized to eYFP controls. * $P < 0.05$, ** $P < 0.01$, by one-way ANOVA; mean + s.e.m. in bar graphs.

is noteworthy given the accumulating evidence that microglia play a role in AD pathology²⁰. Transcriptomic changes after 40 Hz stimulation were positively correlated with changes due to increased neural activity (by NMDA (*N*-methyl-D-aspartate) and bicuculline), and negatively correlated with changes due to silencing activity (by tetrodotoxin) (Extended Data Fig. 3d). The immediate early genes *Nr4a1*, *Arc*, and *Npas4*, which are upregulated by neuronal activity, were elevated as shown by both RNA-seq and reverse transcribed quantitative PCR (RT-qPCR) (Extended Data Fig. 3e).

These transcriptomic results also suggest an engulfing state of microglia. The upregulated genes were positively correlated with gene expression changes induced by macrophage colony-stimulating factor and granulocyte macrophage colony-stimulating factor, known to promote microglial A β uptake²¹ (Extended Data Fig. 3d). RT-qPCR confirmed that upregulated genes included microglial engulfment associated genes *Cd68*, *B2m*, *Bst2*, *Icam1*, and *Lyz2* (Fig. 3c). Microglia-enriched transcriptional regulator *Irf7*, cell adhesion and migration regulator *Spp1*, and microglia proliferation markers *Csf1r* and *Csf2ra* were also upregulated (Fig. 3c). Notably, RT-qPCR showed that the expression levels of pro-inflammatory genes *Il6*, *Il1b* (*Il1- β*), *Itgam* (*Cd11b*), and the anti-inflammatory gene *Igf1* were not changed (Fig. 3c).

Given that 40 Hz stimulation upregulated both phagocytosis- and migration/cell adhesion-related genes, we examined morphological features of microglia activation. We used an antibody against microglial

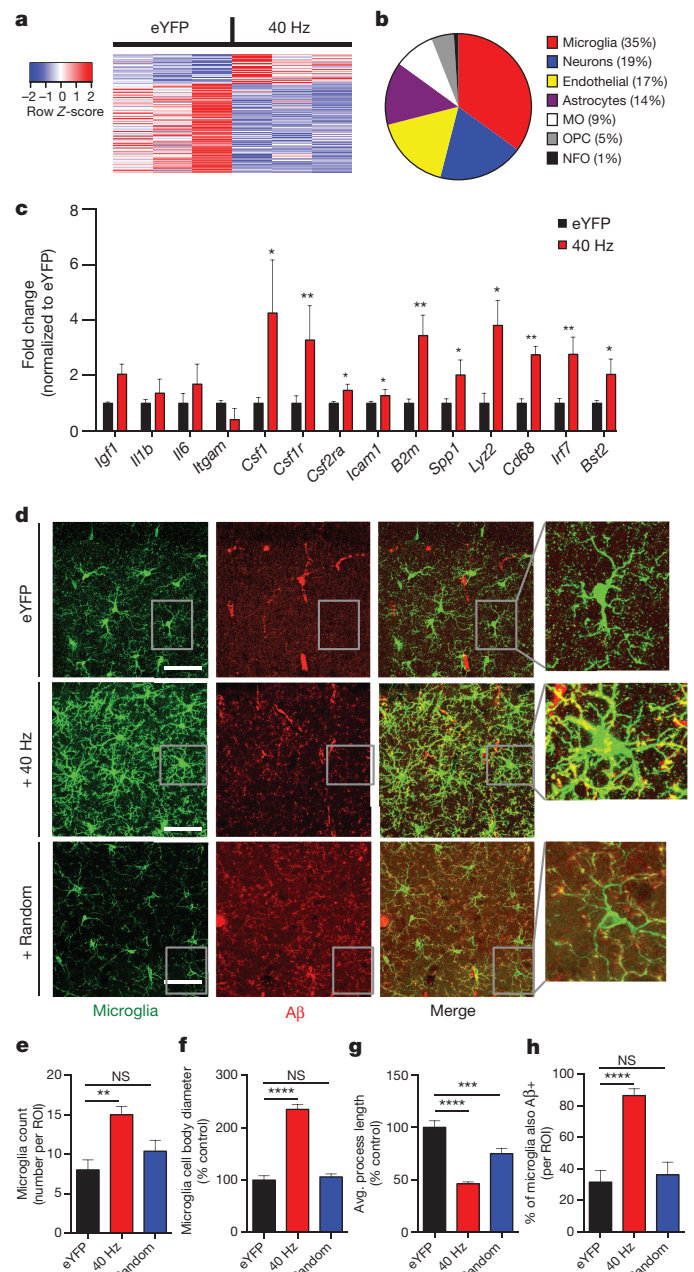


Figure 3 | Driving 40 Hz oscillations optogenetically in hippocampus causes a distinct morphological transformation of microglia in 5XFAD mice. **a**, Heat map of differentially expressed genes determined by whole-transcriptome RNA-Seq of CA1 from 5XFAD/PV-Cre mice expressing only eYFP or Chr2 with 40 Hz stimulation. Normalized z-score values (high: red; low: blue) were calculated for each differentially expressed gene (row). **b**, Cell-type-specific expression patterns of upregulated genes after 40 Hz stimulation (MO, myelinating oligodendrocyte; OPC, oligodendrocyte progenitor cell; NFO, newly formed oligodendrocyte). **c**, RT-qPCR of specific upregulated genes: relative RNA levels (fold change) in CA1 of 5XFAD/PV-Cre mice expressing only eYFP or Chr2 with 40 Hz stimulation, normalized to eYFP controls (Student's *t*-test; $n = 6$ mice per group). **d**, Immunohistochemistry with anti-Iba1 (019-19741, green) to identify microglia and anti-A β (12F4, red) antibodies in CA1 of 5XFAD/PV-Cre mice expressing only eYFP or Chr2 with 40 Hz, and random stimulation (40 \times objective; scale bar, 50 μ m). **e**, Number of Iba1-positive microglia (for **f–i**, one-way ANOVA; $n = 4$ mice per group). **f**, Diameter of Iba1-positive microglia cell bodies. **g**, Average length of Iba1-positive microglia primary processes. **h**, Percentage of Iba1-positive microglia cell bodies that are also A β -positive. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$; mean + s.e.m. in bar graphs.

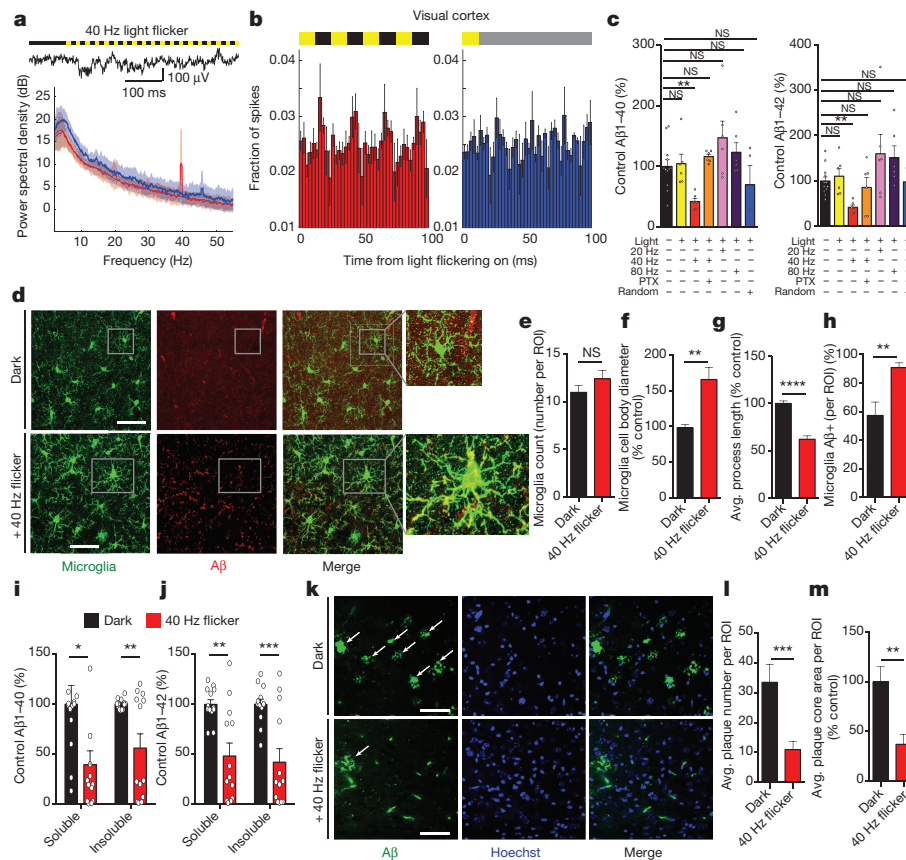


Figure 4 | Driving 40 Hz oscillations in VC via light flicker reduces $A\beta$ and amyloid plaques in 5XFAD mice. **a**, LFP trace in VC before and during 40 Hz light flicker (above). Power spectral density mean and s.d. (below, $n = 4$ 5XFAD mice, five recording sessions). **b**, Fraction of spikes in VC over four cycles of 40 Hz flicker (left) or the equivalent time for random flicker (right, $n = 4$ 5XFAD mice from five recording sessions, mean \pm s.e.m. across animals). For random stimulation, spiking was aligned to light turning on; grey indicates additional light-on flickers occurring randomly (Methods). **c**, Relative $A\beta_{1-40}$ (left) and $A\beta_{1-42}$ (right) levels normalized to dark, in VC of 5XFAD mice exposed to dark, light, 40 Hz, 20 Hz, 80 Hz, 40 Hz with picrotoxin (PTX), and random conditions ($n = 12$ dark; $n = 6$ light, 40 Hz, 20 Hz, 80 Hz flicker and PTX; $n = 4$ random mice; one-way ANOVA). **d**, Immunohistochemistry with anti-Iba1 (019-19741, green) and anti- $A\beta$ (12F4, red) antibodies in VC of

5XFAD mice exposed to dark or 40 Hz flicker. Right: 120 \times zoom; arrows indicate +Iba1/+ $A\beta$ signal in cell body (scale bar, 50 μ m). **e**, Number of Iba1-positive microglia (for **e–h** Student's *t*-test unpaired, $n = 4$ mice per group). **f**, Diameter of Iba1-positive microglia cell bodies. **g**, Average length of Iba1-positive microglia primary processes. **h**, Percentage of Iba1-positive microglia cell bodies that are also $A\beta$ -positive. **i**, Relative $A\beta_{1-40}$ levels in VC of 6-month-old 5XFAD mice after 7 days of 1 h per day dark or 40 Hz flicker (Student's *t*-test unpaired; $n = 13$ mice per group). **j**, As in **i** for $A\beta_{1-42}$. **k**, Immunohistochemistry with anti- $A\beta$ (D5452, green) antibody in 6-month-old VC of 5XFAD mice after 7 days of 1 h per day dark or 40 Hz flicker showing plaques (white arrows; scale bar, 50 μ m). **l**, Number of $A\beta$ -positive plaques; (for **l**, **m**, Student's *t*-test unpaired, $n = 8$ mice per group). **m**, Area of $A\beta$ -positive plaques. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; circles indicate n , mean + s.e.m. in bar graphs.

marker Iba1 to label microglia in CA1 sections from 5XFAD/PV-Cre mice after 1 h of 40 Hz or random stimulation, or in mice expressing only eYFP (Fig. 3d). We observed almost twice as many microglia in the 40 Hz group compared with the eYFP and random control groups (Fig. 3d, e). Furthermore, microglia cell body diameter increased by 135.3% after 40 Hz stimulation compared with eYFP controls and by 138.7% compared with random stimulation (Fig. 3d, f). The lengths of microglia primary processes were reduced by 54.0% in the 40 Hz stimulation condition compared with eYFP controls and by 38.5% compared with random stimulation (Fig. 3d, g). Iba1 levels did not affect these findings, as gene expression analysis showed that Iba1 expression did not differ between conditions (Extended Data Fig. 3a, b). The increase in cell body size and decrease in process length observed after 40 Hz stimulation are consistent with a shift towards a phagocytic state of microglia²². To evaluate microglia $A\beta$ uptake specifically, we measured co-localization of $A\beta$ within microglia by co-immunostaining with an $A\beta$ antibody (12F4, Methods). The percentage of microglia co-localized with $A\beta$ in the cell body increased to 85.6% after 40 Hz stimulation from 31.7% (eYFP control, Fig. 3d, h). Three-dimensional renderings of microglia further demonstrated the presence of $A\beta$ within microglia (Supplementary Videos 1–3). We did not find evidence of neuronal

loss by measuring the CA1 cellular layer thickness (Extended Data Fig. 3f, g). Together, these results suggest that gamma stimulation triggers microglia to increase $A\beta$ uptake.

Light flicker entrains gamma in VC

Many studies have shown that visual stimulation can drive oscillations in the gamma range^{2,23}. In particular, flickering lights at a specific frequency can induce that frequency in primary visual cortex (VC)²⁴. To determine whether light flickering could entrain 40 Hz oscillations to subsequently alter $A\beta$, we exposed 5XFAD mice to 40 Hz flickering for 1 h, analogous to optogenetic stimulation that reduced $A\beta$ described above (Supplementary Video 4 and Methods). In VC, we found that light flickering at 40 Hz increased power in the LFP at 40 Hz, while random interval flickering (random flicker) and dark exposure did not (Fig. 4a and Extended Data Fig. 4a). All induced similar firing rates (Extended Data Fig. 4b, c). Spiking increased and decreased concomitantly as the light flickered on and off, resulting in spiking entrained to 40 Hz during 40 Hz flicker but not during random flicker (Fig. 4b). Recordings from saline above the brain exhibited no change in 40 Hz power during 40 Hz flicker, showing that this effect was not due to photoelectric effects or electrical noise (Extended Data Fig. 4d, e).

Light flicker reduces A β levels in VC

Given that 40 Hz light flicker entrains 40 Hz oscillations in VC, we aimed to determine whether 40 Hz flicker could reduce A β levels. Three-month-old 5XFAD mice were placed in a dark box and exposed to either light flicker at different frequencies (20, 40, or 80 Hz), random flicker, constant light-on (light), or dark for 1 h. One hour after 1 h of 40 Hz flicker, we observed that A β_{1-40} levels in VC were reduced by 57.96% and A β_{1-42} levels by 57.97% compared with dark controls (as measured by A β ELISA, Fig. 4c). The effect was specific to 40 Hz flicker as neither constant light nor 20 Hz, 80 Hz, or random flicker significantly reduced A β levels compared with dark and light controls (Fig. 4c). We also found no change in A β levels in somatosensory barrel cortex and hippocampus after 40 Hz flicker (Extended Data Fig. 5a–h). When we pre-treated 5XFAD mice with a low-dose of GABA_A antagonist (picrotoxin, 0.18 mg/kg (ref. 25)), the effects of 40 Hz flicker on A β levels were completely abrogated, indicating that GABAergic neurotransmission is necessary for this effect (Fig. 4c). To demonstrate that these effects extend beyond the 5XFAD mouse, we examined the effect of 40 Hz flicker in APP/PS1 mice, another well-validated AD model²⁶, and found significantly reduced A β_{1-40} , by 20.80%, and a non-significant trend of reduced A β_{1-42} by 37.68% (Extended Data Fig. 6a). Furthermore, in 9-month-old WT mice, we found a 58.2% reduction in endogenous mouse A β_{1-40} after 1 h of 40 Hz flicker (Extended Data Fig. 6b). The reduction of endogenous mouse A β_{1-40} in WT animals reveals that these effects are not restricted to transgenic APP expression or mutant APP; rather, they extend to A β produced from endogenous APP driven by its endogenous promoter.

Next, we investigated whether 40 Hz flicker alters microglia activity in VC in a similar manner to the way 40 Hz optogenetic stimulation altered CA1 microglia. While microglia number was unchanged (Fig. 4d, e), microglia cell body diameter increased by 165.8% after 40 Hz flicker compared with dark controls (Fig. 4d, f). Microglia primary process lengths were reduced by 37.7% after 40 Hz flicker compared with dark controls (Fig. 4d, g). Consistent with this morphology, which indicates enhanced engulfment activity²², A β /Iba1 co-localization in the cell body was increased to 90.8% after 40 Hz flicker from 57.3% in the dark condition, indicating more A β -bearing microglia (Fig. 4d, h, $P < 0.01$). To better resolve the morphological change in microglia, we used CLARITY²⁷ to create three-dimensional renderings of microglia from 100 μ m sections of VC (Supplementary Videos 5, 6). To demonstrate that microglia indeed engulf A β in the 5XFAD mouse, we purified microglia from 5XFAD and WT animals using fluorescence-activated cell sorting (FACS) and analysed A β levels via human ELISA. We found that microglia-specific levels of A β are significantly higher in 5XFAD animals than in WT controls (Extended Data Fig. 7a, b). Synaptophysin levels did not change between dark and 40 Hz flicker conditions, indicating that microglia activation probably did not affect synapse number (Extended Data Fig. 7c, d; NS, not significant, $n = 4$ mice). Thus, 40 Hz oscillations induced non-invasively via sensory entrainment reduced A β abundance and promoted microglia/A β interactions.

Light flicker decreases plaque load in VC

We next assessed whether 40 Hz flicker was effective in treating animals that have amyloid plaques. Because the effects of 40 Hz flicker on soluble A β levels were transient, lasting 12 h but not 24 h (Extended Data Fig. 8a), we hypothesized that we would need repeated exposure to affect insoluble A β . Thus, we treated 6-month-old 5XFAD mice, which have amyloid plaque pathology in many brain regions including VC¹⁰, for 1 h daily over 7 days with 40 Hz flicker or dark, and analysed VC tissue 24 h later. We found that 7 days of 1 h 40 Hz flicker reduced both soluble A β_{1-40} and A β_{1-42} levels, by 60.5% and 51.7% respectively, (Fig. 4i, j) and insoluble A β_{1-40} and A β_{1-42} levels by 43.7% and 57.9% respectively (Fig. 4i, j). Immunohistochemical analysis showed that 40 Hz flicker significantly reduced plaque number in VC by 67.2% compared with dark controls (Fig. 4k, l) and plaque size by 63.7%

(Fig. 4k, m). Taken together, these experiments identify a non-invasive treatment with a profound effect on amyloid plaque pathology. We next determined if 40 Hz flicker reduced tau phosphorylation by immunohistochemistry, another AD-related pathology. Using the TauP301S tauopathy mouse model²⁸ we found that 7 days of 1 h 40 Hz visual flicker treatment reduced phosphorylated tau serine202 and serine404/threonine403/serine400 levels in VC by 41.2% and 42.3%, respectively, and triggered microglia responses similar to those observed in 5XFAD mice (Extended Data Fig. 9a–k).

Discussion

Gamma oscillations are thought to be important for higher cognitive functions and sensory responses^{2,12,23}. Here, we demonstrated that entraining oscillations and spiking at 40 Hz, using optogenetics in the hippocampus of 5XFAD mice and using a non-invasive light flicker treatment to affect primary VC in multiple mouse models, resulted in a marked reduction of A β peptides. We also found a concomitant microglia response after 40 Hz entrainment.

The robust reduction of total amyloid levels was probably mediated both by decreased amyloidogenesis and by increased amyloid endocytosis by microglia. Thus, it appears that driving 40 Hz gamma oscillations may induce an overall neuroprotective response that recruits both neurons and microglia. The fact that GABA_A antagonist treatment completely abrogated the effects of 40 Hz stimulation on A β levels strongly suggests that GABAergic neurotransmission is critical for these effects.

Flicker stimulation at 40 Hz reduced A β in multiple mouse models, including 5XFAD, APP/PS1, and WT mice. This replication in multiple mouse models shows that these findings are not specific to one animal model and, importantly, extend to situations where A β is produced from APP expressed by its physiological promoter as it is in WT animals. In addition, we found that 40 Hz oscillations reduced phosphorylated tau staining in a mouse model of tauopathy, TauP301S, showing that the protective effects of gamma stimulation generalize to other pathogenic proteins.

These observations indicate that entraining gamma oscillations may provide a broad spectrum of systemic effects in the brain, including in non-neuronal cells, to attenuate AD-related pathology. Because this approach is fundamentally different from previous AD therapies²⁹, further study is needed to determine whether it will be therapeutic in human AD.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 February; accepted 2 November 2016.

- Bartos, M., Vida, I. & Jonas, P. Synaptic mechanisms of synchronized gamma oscillations in inhibitory interneuron networks. *Nature Rev. Neurosci.* **8**, 45–56 (2007).
- Fries, P., Nikolic, D. & Singer, W. The gamma cycle. *Trends Neurosci.* **30**, 309–316 (2007).
- Cardin, J. A. *et al.* Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* **459**, 663–667 (2009).
- Palop, J. J. *et al.* Aberrant excitatory neuronal activity and compensatory remodeling of inhibitory hippocampal circuits in mouse models of Alzheimer's disease. *Neuron* **55**, 697–711 (2007).
- Verret, L. *et al.* Inhibitory interneuron deficit links altered network activity and cognitive dysfunction in Alzheimer model. *Cell* **149**, 708–721 (2012).
- Bero, A. W. *et al.* Neuronal activity regulates the regional vulnerability to amyloid- β deposition. *Nature Neurosci.* **14**, 750–756 (2011).
- Selkoe, D. J. *et al.* The role of APP processing and trafficking pathways in the formation of amyloid beta-protein. *Ann. NY Acad. Sci.* **777**, 57–64 (1996).
- Stam, C. J. *et al.* Generalized synchronization of MEG recordings in Alzheimer's disease: evidence for involvement of the gamma band. *J. Clin. Neurophysiol.* **19**, 562–574 (2002).
- Gillespie, A. K. *et al.* Apolipoprotein E4 causes age-dependent disruption of slow gamma oscillations during hippocampal sharp-wave ripples. *Neuron* **90**, 740–751 (2016).
- Oakley, H. *et al.* Intraneuronal β -amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J. Neurosci.* **26**, 10129–10140 (2006).

11. Colgin, L. L. *et al.* Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature* **462**, 353–357 (2009).
12. Buzsáki, G. *Rhythms of the Brain* (Oxford Univ. Press, 2006).
13. Buzsáki, G. *et al.* Hippocampal network patterns of activity in the mouse. *Neuroscience* **116**, 201–211 (2003).
14. Carr, M. F., Karlsson, M. P. & Frank, L. M. Transient slow gamma synchrony underlies hippocampal memory replay. *Neuron* **75**, 700–713 (2012).
15. Foster, D. J. & Wilson, M. A. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440**, 680–683 (2006).
16. Carr, M. F., Jadhav, S. P. & Frank, L. M. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature Neurosci.* **14**, 147–153 (2011).
17. Helwig, M. *et al.* The neuroendocrine protein 7B2 suppresses the aggregation of neurodegenerative disease-related proteins. *J. Biol. Chem.* **288**, 1114–1124 (2013).
18. Das, U. *et al.* Activity-induced convergence of APP and BACE-1 in acidic microdomains via an endocytosis-dependent pathway. *Neuron* **79**, 447–460 (2013).
19. Cataldo, A. M. *et al.* Endocytic pathway abnormalities precede amyloid beta deposition in sporadic Alzheimer's disease and Down syndrome: differential effects of APOE genotype and presenilin mutations. *Am. J. Pathol.* **157**, 277–286 (2000).
20. Gjonneska, E. *et al.* Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* **518**, 365–369 (2015).
21. Chitu, V. & Stanley, E. R. Colony-stimulating factor-1 in immunity and inflammation. *Curr. Opin. Immunol.* **18**, 39–48 (2006).
22. Wang, Y. *et al.* TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. *Cell* **160**, 1061–1071 (2015).
23. Gray, C. M., König, P., Engel, A. K. & Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338**, 334–337 (1989).
24. Eckhorn, R. *et al.* Coherent oscillations: a mechanism of feature linking in the visual cortex? Multiple electrode and correlation analyses in the cat. *Biol. Cybern.* **60**, 121–130 (1988).
25. Perićić, D. & Bujas, M. Sex differences in the response to GABA antagonists depend on the route of drug administration. *Exp. Brain Res.* **115**, 187–190 (1997).
26. Jankowsky, J. L. *et al.* Mutant presenilins specifically elevate the levels of the 42 residue β -amyloid peptide *in vivo*: evidence for augmentation of a 42-specific γ secretase. *Hum. Mol. Genet.* **13**, 159–170 (2004).
27. Chung, K. *et al.* Structural and molecular interrogation of intact biological systems. *Nature* **497**, 332–337 (2013).
28. Yoshiyama, Y. *et al.* Synapse loss and microglial activation precede tangles in a P301S tauopathy mouse model. *Neuron* **53**, 337–351 (2007).
29. Canter, R. G., Penney, J. & Tsai, L.-H. The road to restoring neural circuits for the treatment of Alzheimer's disease. *Nature* **539**, 187–196 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to S. Tonegawa and D. Roy for APP/PS1 mice, and E. Demmons, W. Raja, E. Wu, and B. Arkhurst and the Boyden laboratory for technical assistance. We thank members of the Tsai and Boyden laboratories, C. Moore, C. Deister, D. Rei, J. Penney, R. Madabhushi, A. Mungenast, A. Bero, and J. Young for discussions and comments on the paper. H.F.I. acknowledges the Cameron Hayden Lord Foundation and Barbara J. Weedon Fellowship; E.S.B. acknowledges the New York Stem Cell Foundation-Robertson Award, National Institutes of Health (NIH) 1R01EY023173, and NIH 1DP1NS087724; L.H.T. acknowledges the JPB Foundation, Belfer Neurodegeneration Consortium, Halis Family Foundation, and NIH RF1 AG047661. E.S.B. and E.N.B. acknowledge NIH ROI GM104948.

Author Contributions H.F.I., A.C.S., E.N.B., E.S.B., and L.-H.T. designed experiments. H.F.I. and F.G. performed RNA sequencing experiments. H.F.I. and A.C.S. performed electrophysiology. A.C.S. analysed electrophysiology data. H.F.I. performed and analysed optogenetics and ELISA experiments. T.Z.G., J.S., and O.K. performed western blots. H.F.I., A.R., F.A., R.R., and R.G.C. performed and analysed imaging experiments. F.G. analysed RNA sequencing data. H.F.I., A.J.M., and C.A. performed visual stimulation experiments. H.M. performed FACS experiments. H.F.I., A.C.S., A.R., F.G., E.S.B., and L.-H.T. wrote the manuscript.

Author Information RNA-seq data available at Gene Expression Omnibus under accession number GSE77471. Other data are publicly available upon request. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.-H.T. (lhtsai@mit.edu).

METHODS

Animals. All animal work was approved by the Committee for Animal Care of the Division of Comparative Medicine at the Massachusetts Institute of Technology. Adult (3-month-old) male double transgenic 5XFAD Cre mice were produced by crossing 5XFAD transgenic mice with the transgenic PV or CW2 promoter driven Cre line. Adult (5-month-old) male and female APP/PS1 mice were gifted from the Tonegawa laboratory. Adult (4-month-old) male TauP301S mice were obtained from the Jackson Laboratory. Nine-month-old WT mice (C57Bl/6) were obtained from the Jackson Laboratory. Mice were housed in groups of three to five on a standard 12 h light/12 h dark cycle, and all experiments were performed during the light cycle. Food and water were provided *ad libitum* unless otherwise noted. Littermates were randomly assigned to each condition by the experimenter. The experimenter was blind to animal genotypes during tissue processing and electrophysiological recording and analysis. No animals were excluded from analysis.

Adeno-associated viral vectors. Adeno-associated viral (AAV) particles of serotype 5 were obtained from the Vector Core Facility at The University of North Carolina at Chapel Hill. The AAV5 virus contained a channelrhodopsin-2 (ChR2) fused to eYFP in a double-floxed, inverted, open-reading-frame (DIO) driven by the EF1 α promoter (Extended Data Fig. 2a). An AAV-DIO-eYFP construct was used as a control.

Surgical procedures. Three-month-old 5XFAD/PV-Cre or CW2 mice were anaesthetized with an intraperitoneal (i.p.) injection of a mixture of ketamine (1.1 mg/kg) and xylazine (0.16 mg/kg). A small craniotomy was made 2.0 mm posterior to bregma and 1.8 mm lateral to the midline on the left side. Virus was delivered through a small durotomy by a glass micropipette attached to a Quintessential Stereotaxic Injector (Stoelting). The glass micropipette was lowered to 1.2 mm below the brain surface. A bolus of 1 μ L of virus (AAV-DIO-ChR2-eYFP or AAV-DIO-eYFP; 2×10^{12} viral molecules per millilitre) was injected into the CA1 region of the hippocampus at $0.075 \mu\text{L min}^{-1}$. The pipette remained in place for 5 min following the injection before being retracted from the brain. A unilateral optical fibre implant (300 μ m core diameter; Thor Labs) was lowered to 0.9 mm below the brain surface about the injection site. Two small screws anchored at the anterior and posterior edges of the surgical site were bound with dental glue to secure the implant in place.

For electrophysiological recordings, adult (3-month-old) male 5XFAD/PV-Cre and 5XFAD negative littermates (for CA1 recordings), or 5XFAD and their WT littermates (for VC recordings) mice were anaesthetized using isoflurane and placed in a stereotaxic frame. The scalp was shaved, ophthalmic ointment (Puralube Vet Ointment, Dechra) was applied to the eyes, and Betadine and 70% ethanol were used to sterilize the surgical area. For CA1 recordings, a craniotomy (in millimetres, from bregma: -2 anterior/posterior, 1.8 medial/lateral) was opened to deliver 1 μ L of virus to CA1 (as described above). The target craniotomy site for LFP recordings was marked on the skull (in mm, from bregma: -3.23 anterior/posterior, 0.98 medial/lateral for CA1 and 2.8 anterior/posterior, 2.5 medial/lateral for VC), three self-tapping screws (F000CE094, Morris Precision Screws and Parts) were attached to the skull, and a custom stainless steel headplate was affixed using dental cement (C&B Metabond, Parkell). On the day of the first recording session, a dental drill was used to open the LFP craniotomies (300–400 μ m diameter) by first thinning the skull until $\sim 100 \mu\text{m}$ thick, and then using a 30-gauge needle to make a small aperture. The craniotomy was then sealed with a sterile silicone elastomer (Kwik-Sil WPI) until recording that day and in between recording sessions.

Optogenetic stimulation protocol. Two to four weeks after virus injection and implant placement (which provided time for the mice to recover and undergo behaviour training for animals used for electrophysiology, and the virus to express in the neurons), CA1 neurons were optogenetically manipulated. A 200 mW, 4,793 nm DPSS laser was connected to a patch cord with a fibre channel/physical contact connector at each end. During the experiment, 1 mW (measured from the end of the fibre) of optical stimulation was delivered for 1 h. For molecular and biochemical analyses, each animal received one of three stimulation protocols: 8 Hz, 40 Hz, or random stimulation (light pulses were delivered with a random interval determined by a Poisson process with an average frequency of 40 Hz). eYFP control animals received 40 Hz stimulation. For electrophysiological recordings, each animal received all stimulation conditions interleaved during recordings.

Visual stimulation protocol. Fifteen minutes before the experiment, 5XFAD mice were treated with saline (control) or picrotoxin (0.18 mg/kg)²⁵. For molecular and biochemical analyses, mice were then placed in a dark chamber illuminated by a light-emitting diode (LED) bulb and exposed to one of five stimulation conditions: dark, light, 20 Hz, 40 Hz (12.5 ms light on, 12.5 ms light off, 60 W), 80 Hz flicker for 1 h. For electrophysiological recordings, each animal received dark, light, 40 Hz flicker, or random (light pulses were delivered with a random interval determined by a Poisson process with an average interval of 40 Hz) stimulation conditions interleaved in 10 s blocks during recordings.

Behaviour training and virtual reality environment (VR) for electrophysiology.

For CA1 recordings, head-fixed animals ran on an 8-inch spherical treadmill supported by an air cushion through a virtual reality environment, as described in ref. 30. The motion of the spherical treadmill was measured by an optical mouse and fed into virtual reality software³¹, running in MATLAB (version 2013b, Mathworks). The virtual environment consisted of a linear track with two small enclosures at the ends where the animal could turn (Extended Data Fig. 1a). Animals were rewarded with sweetened condensed milk (diluted 1:2 in water) at each end of the track for alternating visits to each end of the track.

Animals learned to run on the virtual linear track over approximately 1 week. The animals were left to recover from the surgery for 1 week, and habituated to handling for 1–2 days before behavioural training began. To learn to manoeuvre on the treadmill and get comfortable in the testing environment, on the first 2 days of training the animals were placed on the spherical treadmill with the virtual reality system off and were rewarded with undiluted sweetened condensed milk. On the second day of training on the spherical treadmill, the animals' food was restricted to motivate them to run. Animals were restricted to no more than 85% of their baseline weight and typically weighed over 88% of their baseline weight. From the third day until the end of training (typically 5–7 days) the animals were placed on the treadmill for increasing amounts of time (30 min to 2 h) running in the VR linear track. Animals were rewarded with diluted (1:2) sweetened condensed milk at the end of the linear track after traversing the length of the track. Between recording sessions, animals were given refresher training sessions to maintain behavioural performance.

For VC recordings, animals ran on the spherical treadmill while exposed to dark, light, or light-flickering conditions (described below in data acquisition). Before recordings, animals learned to manoeuvre on the treadmill and get comfortable in the testing environment by being placed on the spherical treadmill (with the virtual reality system off) and receiving a reward of undiluted sweetened condensed milk.

Electrophysiology data acquisition. For optogenetic stimulation of CA1 during recording, a 300 μ m core optical fibre was advanced through the craniotomy used to deliver virus to CA1 to a depth of 900 μ m into the brain. Light pulses that were 1 ms and 1 mW (measured from the end of the fibre) were delivered via a 473 nm DPSS (diode pumped solid state) laser (as described above).

To avoid photoelectric artefacts, neural activity was recorded with glass electrodes. LFP electrodes were pulled from borosilicate glass pipettes (Warner) on a filament-based micropipette puller (Flaming-Brown P97, Sutter Instruments), to a fine tip, which was then manually broken back to a diameter of ~ 10 – $20 \mu\text{m}$ and filled with sterile saline. For CA1 recordings the LFP electrode was advanced through the LFP recording craniotomy at an angle 60 degrees posterior to the coronal plane and 45° inferior to the horizontal plane until clear electrophysiological signatures of the hippocampal stratum pyramidale layer were observed (~ 600 – $1000 \mu\text{V}$ theta waves while the animal was running, clearly distinguishable SWRs during immobility, and multiple spikes greater than $150 \mu\text{V}$; Extended Data Fig. 1b). For VC recordings, the LFP electrode was advanced vertically through the LFP recording craniotomy to a depth of 600–900 μm and multiple spikes greater than $150 \mu\text{V}$ were observed.

Data were acquired with a sampling rate of 20 kHz and bandpass filtered 1 Hz to 1 kHz. Animals ran on the spherical treadmill or rested for prolonged periods. For optogenetic stimulation sessions, data were recorded for 30 min before any stimulation began. Then stimulation was delivered at gamma (40 Hz), random (as described under Optogenetic stimulation protocol), or theta (8 Hz) frequency for 10 s periods interleaved with 10 s baseline periods (no stimulation). In two animals, stimulation of each type or baseline was delivered for 5 min periods instead of 10 s periods. Each 30 min of stimulation recordings were followed by 5–30 min of recording with no stimulation. For visual light flicker stimulation sessions, LED strip lights surrounding the animal lights were flickered at gamma (40 Hz), random (described above in Visual stimulation protocol), theta (8 Hz), or 20 Hz frequency for 10 s periods, or were on continuously for 10 s periods, interleaved with 10 s periods with lights off. A few recordings were made above the brain surface during light flicker to ensure that the lights did not create electrical or photoelectric noise during recording. Recording sessions were terminated after approximately 3–5 h. Animals were 3–4 months old at the time of recording.

Spike detection. Spikes were detected by thresholding the 300–6,000 Hz band-passed signal. Threshold was the median of the filtered signal plus five times a robust estimator of the standard deviation of the filtered signal (median/0.675) to avoid contamination of the standard deviation measure by spikes³².

LFP. Recorded traces were downsampled to 2 kHz and then bandpass filtered between 1 and 300 Hz.

Theta and SWR detection. Activity across the hippocampal network changes markedly when animals run or sit quietly, and these changes are often referred to

as different network states. These network states are clearly distinguishable by the presence or absence of LFP oscillations in different frequency bands^{12,13}. When animals ran, we observed large theta (4–12 Hz) oscillations in CA1 as others have shown (Extended Data Fig. 1b, left)^{13,30,33,34}. When animals sat quietly, theta oscillations were no longer visible and we recorded SWRs, high-frequency oscillations of 150–250 Hz that last around 50–100 ms and are associated with bursts of population activity, as others have observed (Extended Data Fig. 1b, right)^{15,16}.

SWRs were detected (Fig. 1a–d and Extended Data Fig. 1d–i) when the envelope amplitude of the filtered trace was greater than four standard deviations above the mean for at least 15 ms. The envelope amplitude was calculated by taking the absolute value of the Hilbert transform of the filtered LFP. We also confirmed our results held when using a higher threshold for SWR detection, six standard deviations above the mean, which detects larger SWRs (Extended Data Fig. 1j, k). To detect theta (Extended Data Fig. 1c, d), the LFP was bandpass filtered for theta (4–12 Hz), delta (1–4 Hz), and beta (12–30 Hz) using an FIR equiripple filter. The ratio of theta to delta and beta ('theta ratio') was computed as the theta envelope amplitude divided by the sum of the delta and beta envelope amplitudes. Theta periods were classified as such when the theta ratio was greater than one standard deviation above mean for at least 2 s and the ratio reached a peak of at least two standard deviations above mean. Non-theta periods were classified as such when the theta ratio was less than one for at least 2 s. SWRs, theta periods, and non-theta periods were visually inspected to ensure that these criteria accurately detected SWRs, theta periods, and non-theta periods, respectively.

Power spectrum. Spectral analysis was performing using multitaper methods (Chronux toolbox, time-bandwidth product = 3, number of tapers = 5). For examining power spectra without stimulation (Extended Data Fig. 1c, d), only theta periods were included: theta periods greater than 5 s long were divided into 5 s trials and the average power spectral density was computed for each animal over these trials. For examining power spectra during optogenetic (Fig. 1e and Extended Data Fig. 1l) and visual stimulation (Fig. 4a and Extended Data Fig. 4a), data were divided into 10 s trials of each stimulation condition or baseline periods, and the average power spectral density was computed for each animal over these trials.

Gamma during SWRs. Spectrograms were computed using multitaper methods (Chronux toolbox). The spectrogram was computed for each SWR including a window of 400 ms before and after the peak of the SWR. Then a z-scored spectrogram was computed in each frequency band using the mean and standard deviation of the spectrogram computed across the entire recording session to create a normalized measure of power in units of standard deviation (Fig. 1a and Extended Data Fig. 1e). Instantaneous frequency of gamma during SWRs was computed by bandpass filtering the LFP for 10–50 Hz, taking the Hilbert transform, then taking the reciprocal of the difference in peaks of the transformed signal (Fig. 1b and Extended Data Fig. 1f). Gamma power before, during, and after SWRs was computed by filtering the LFP for low gamma (20–50 Hz) and taking the amplitude of the envelope of the Hilbert transform to get the mean gamma power in 100 ms bins centred on the SWR peak. This was normalized by the mean and standard deviation of the amplitude of the envelope for the entire recording session to get z-scored gamma power for each bin around each SWRs (Fig. 1c and Extended Data Fig. 1g, j). Phase modulation by gamma during SWRs was computed by bandpass filtering the LFP for gamma (20–50 Hz), taking the Hilbert transform, and determining the phase of the resulting signal for each spike that occurred during SWRs (Extended Data Fig. 1h). To measure differences in phase modulation between 5XFAD and WT animals, we used resampling with replacement: a subset of 100 spikes from each recording was randomly selected to create a phase modulation distribution and this was repeated 500 times for each recording (Fig. 1d and Extended Data Fig. 1k). We then measured the depth of modulation for the spike-gamma phase distribution by computing the difference between the peak and trough divided by the sum of the peak and trough for each distribution (Fig. 1d and Extended Data Fig. 1k).

Differences in firing during stimulation. To plot stimulus-evoked multiunit firing histograms, spikes were binned in 2.5 ms bins for 100 ms after the start of each light-on pulse and the fraction of spikes in each bin was computed. Mean and standard error were then computed across all light-on periods. To compute differences in multi-unit firing rate between conditions, firing rates were computed for each 10 s period of stimulation or baseline (total number of spikes divided by duration of period). Differences in firing rate were taken between nearby periods of the relevant type of stimulation (firing rate in gamma stimulation period minus baseline or random periods for optogenetic stimulation, firing rate in gamma stimulation period minus baseline, continuous on, or random periods for light flicker stimulation). Differences from all animals were plotted in histograms (Extended Data Figs 1m and 4c) and the median and quartiles of the multiunit firing rates per 40 Hz stimulation, random stimulation, and no stimulation period for each animal were plotted in box plots (Extended Data Figs 1o and 4d).

Immunohistochemistry. Mice were perfused with 4% paraformaldehyde under deep anaesthesia, and the brains were post-fixed overnight in 4% paraformaldehyde. Brains were sectioned at 40 μ m using a vibratome (Leica). Sections were permeabilized and blocked in PBS containing 0.2% Triton X-100 and 10% normal donkey serum at room temperature for 1 h. Sections were incubated overnight at 4°C in primary antibody in PBS with 0.2% Triton X-100 and 10% normal donkey serum. Primary antibodies were anti-EEA1 (BD Transduction Laboratories; 641057), anti- β -amyloid (Cell Signaling Technology; D54D2), anti-Iba1 (Wako Chemicals; 019-19741), anti-parvalbumin (Abcam; ab32895), and anti-Rab5 (Enzo Life Sciences; ADI-KAP-GP006-E). To confirm ELISA experiments, the anti-A β antibody D54D2 was used because it allowed for co-labelling with EEA1 and the anti-A β antibody 12F4 was used because it does not react with APP, allowing us to determine whether our labelling was specific to A β . For co-labelling experiments, the anti-A β antibody 12F4 (Biolegend; 805501) was used. Primary antibodies were visualized with Alexa-Fluor 488 and Alex-Fluor 647 secondary antibodies (Molecular Probes), and cell nuclei visualized with Hoechst 33342 (Sigma-Aldrich; 94403). Images were acquired using a confocal microscope (LSM 710; Zeiss) with a 40 \times objective at identical settings for all conditions. Images were quantified using ImageJ 1.42q by an experimenter blind to treatment groups. For each experimental condition, two coronal sections from at least three animals were used for quantification. Scale bars are 50 μ m. For CA1 imaging, the analysis was restricted to the pyramidal cell layer, except in the case of Iba1 + cells analysis, where the whole field of view was required to image an adequate number of cells. ImageJ was used to measure the diameter of Iba1 + cell bodies and to trace the processes for length measurement. In addition, the Coloc2 plugin was used to measure co-localization of Iba1 and A β . Imarisx64 8.1.2 (Bitplane, Zurich, Switzerland) was used for three-dimensional rendering. For counting the 'plaque number', deposits of at least 10 μ m were included.

CLARITY. Fixed brains were sliced into 100 μ m coronal sections on a vibratome (Leica VT100S) in 1 \times PBS. Sections containing VC were selected, with reference to the Allen Mouse Brain Atlas, and incubated in clearing buffer (pH 8.5–9.0, 200 mM sodium dodecylsulfate, 20 mM lithium hydroxide monohydrate, 4 mM boric acid in double-distilled H₂O) for 2 h, shaking at 55°C. Cleared sections were washed 3 \times 10 mins in 1 \times PBST (0.1% Triton-X100/1XPBS) and put into blocking solution (2% bovine serum albumin/1 \times PBST) overnight, shaking at room temperature²⁷. Subsequently, three 1 h washes in 1 \times PBST were performed, shaking at room temperature. Sections were then incubated at 4°C for 2 days, shaking, with anti- β -amyloid (Biolegend; 805501) and anti-Iba1 (Wako Chemicals; 019-19741) primary antibodies, diluted to 1:100 in 1 \times PBST. Another set of 3 \times 1 h washes in 1 \times PBST was conducted before sections were incubated for 9 h, shaking at room temperature, in 1:100 1 \times PBS-diluted secondary antibody mixture. Fragmented Donkey Anti-Rabbit Alexa Fluor 488 (Abcam; ab175694) and Anti-Mouse 568 (Abcam; ab150101) secondary antibodies were used to visualize the primary antibody labelling. Halfway through this incubation period, Hoechst 33258 (Sigma-Aldrich; 94403) was spiked into each sample at a 1:250 final dilution. Sections were then washed overnight in 1 \times PBS, shaking at room temperature. Before mounting for imaging, slices were incubated in refractive index matching solution (75 g Histodenz, 20 mL 0.1 M phosphate buffer, 60 mL double-distilled H₂O) for 1 h, shaking at room temperature. Tissue sections were mounted onto microscopy slides with coverslips (VWR VistaVision, VWR International, LLC, Radnor, Pennsylvania, USA) using Fluoromount G Mounting Medium (Electron Microscopy Sciences, Hatfield, Pennsylvania, USA). Images were acquired on a Zeiss LSM 880 microscope with the accompanying Zen Black 2.1 software (Carl Zeiss Microscopy, Jena, Germany). Section overview and cellular-level images used for three-dimensional reconstruction were taken using a Plan-Apochromat 63 \times /1.4 oil differential interference contrast objective. Imarisx64 8.1.2 (Bitplane, Zurich, Switzerland) was used for three-dimensional rendering and analysis.

Western blot. CA1 whole-cell lysates were prepared using tissue from 3-month-old male 5XFAD/PV-Cre mice. Tissue was homogenized in 1 ml RIPA (50 mM Tris HCl pH 8.0, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) buffer with a hand homogenizer (Sigma), incubated on ice for 15 min, and rotated at 4°C for 30 min. Cell debris was isolated and discarded by centrifugation at 14,000 r.p.m. for 10 min. Lysates were quantitated using a nanodrop, and 25 μ g protein was loaded on a 10% acrylamide gels. Protein was transferred from acrylamide gels to PVDF membranes (Invitrogen) at 100 V for 120 min. Membranes were blocked using bovine serum albumin (5% w/v) diluted in TBS:Tween. Membranes were incubated in primary antibodies overnight at 4°C and secondary antibodies at room temperature for 90 min. Primary antibodies were anti-APP (Invitrogen; PAD CT695), anti-APP (Sigma; A8967), and anti- β -actin (Abcam; ab9485). Secondary antibodies were horseradish peroxidase-linked (GE Healthcare). Signal intensities were quantified using ImageJ 1.46a and normalized to values of β -actin. We examined tau protein solubility using sequential protein extraction as described in

ref. 28. We then probed the detergent insoluble tau fraction using an antibody against Tau5 (Thermo Fisher Scientific; AHB0042).

ELISA. CA1 or VC was isolated from male mice, lysed with PBS or 5M Guanidine HCl, and subjected to A β measurement with the use of mouse (for WT experiments) or human (for all other experiments) A β _{1–40} or A β _{1–42} ELISA kit (Invitrogen) according to the manufacturer's instructions. We lysed the tissue in phosphate-buffered saline (PBS) to extract the PBS soluble A β fraction. The soluble A β fraction probably contained monomeric and oligomeric A β . Tissue was further treated with guanidine HCl to extract the insoluble A β fraction. A β _{1–42} was below detectable levels for both flicker and control groups in WT VC and microglia-specific samples.

Genome-wide RNA-seq. Total RNA was extracted from CA1 isolates using the RNeasy kit (Qiagen). Purified mRNA was used for RNA-seq library preparation using the BIONEXTflex kit (BIONEXT 5138-08) according to the manufacturer's instructions. Total mRNA (1 μ g) was subject to a sequential workflow of poly-A purification, fragmentation, first strand and second strand synthesis, DNA end-adenylation, and adaptor ligation. The libraries were enriched by 15 cycles of PCR reactions and cleaned with Agencourt AMPure XP magnetic beads (Beckman Coulter). The quality of the libraries was assessed using an Advanced Analytical-fragment Analyzer. The bar-coded libraries were equally mixed for sequencing in a single lane on the Illumina HiSeq 2000 platform at the MIT BioMicro Center. The raw fastq data of 50-bp single-end sequencing reads were aligned to the mouse mm9 reference genome using TopHat2.0. The mapped reads were processed by Cufflinks 2.2 with UCSC mm9 reference gene annotation to estimate transcript abundances, and test for differential expression. An average of 26,518,345 sequencing reads was obtained from three stimulated and three non-stimulated mice. Relative abundance of transcript was measured by fragments per kilobase of exon per million fragments mapped (FPKM). Gene differential expression test between treated and untreated groups was performed using Cuffdiff module with an adjusted *P* value <0.05 for statistical significance (GEO accession number GSE77471).

To understand the cellular and molecular mechanisms from our RNA-seq data, 14 of publicly available RNA-seq datasets³⁵ were processed for cell-type-specific analysis. Additionally, 60 publicly available neuron-, microglia-, and macrophage-specific RNA-seq datasets under different chemical and genetic perturbations^{36–41} were downloaded and processed using TopHat/Cufflinks pipeline for gene set enrichment (GSEA) statistical analysis. GSEA was used to determine whether a defined gene set from our RNA-seq data are significantly enriched at either direction of a ranked gene list from a particular perturbation study. Genes detected in the public RNA-seq datasets were ranked by log₂ values of fold change (case versus control), from positive to negative values. A defined gene set (in our case, up- or downregulated genes upon gamma treatment) was considered significantly correlated with a perturbation-induced transcriptomic change (either up- or down-regulation) when both nominal *P* value and false discovery rate *q* value were less than 0.05. The sign of the calculated normalized enrichment score (NES) indicates whether the gene set is enriched at the top or the bottom of the ranked list. The heatmap for differentially expressed genes was generated using a custom R script, and *z*-score values across all libraries for each gene were calculated on the basis of the gene FPKM values. The box plots for cell-type specificity analysis were also generated by the R program, on the basis of gene FPKM values.

Quantitative RT–PCR. The CA1 subregion was isolated from hippocampus of 3-month-old male 5XFAD/PV-Cre mice. Tissue was rapidly frozen using liquid nitrogen and stored at –80 °C, and RNA extracted using the RNeasy kit according to the manufacturer's protocol (Qiagen). RNA (3 μ g) was treated with DNase I (4 U, Worthington Biochemical Corporation), purified using RNA Clean and Concentrator-5 Kit (Zymo Research) according to the manufacturers' instructions, and eluted with 14 μ L DEPC-treated water. For each sample, 1 μ g RNA was reverse transcribed in a 20 μ L reaction volume containing random hexamer mix and Superscript III reverse transcriptase (50 U, Invitrogen) at 50 °C for 1 h. First strand cDNAs were diluted 1:10 and 1 μ L were used for RT-qPCR amplification in a 20 μ L reaction (SsoFast EvaGreen Supermix, Bio-Rad) containing primers (0.2 μ M). Relative changes in gene expression were assessed using the 2^{– $\Delta\Delta C_t$} method.

Isolation of microglia from VC. The primary VC (V1 region) was rapidly dissected and placed in ice-cold Hanks' balanced salt solution (HBSS) (Gibco by

Life Technologies, catalogue number 14175-095). The tissue was then enzymatically digested using the Neural Tissue Dissociation Kit (P) (Miltenyi Biotec, catalogue number 130-092-628) according to the manufacturer's protocol, with minor modifications. Specifically, the tissue was enzymatically digested at 37 °C for 15 min instead of 35 min and the resulting cell suspension was passed through a 40 μ m cell strainer (Falcon Cell Strainers, Sterile, Corning, product 352340) instead of a MACS SmartStrainer, 70 μ m. The resulting cell suspension was then stained using allophycocyanin (APC)-conjugated CD11b mouse clone M1/70.15.11.5 (Miltenyi Biotec, 130-098-088) and phycoerythrin (PE)-conjugated CD45 antibody (BD Pharmingen, 553081) according to the manufacturer's (Miltenyi Biotec) recommendations. FACS was then used to purify CD11b and CD45 positive microglial cells. The cells were sorted directly into 1 \times PBS (Extended Data Fig. 6a).

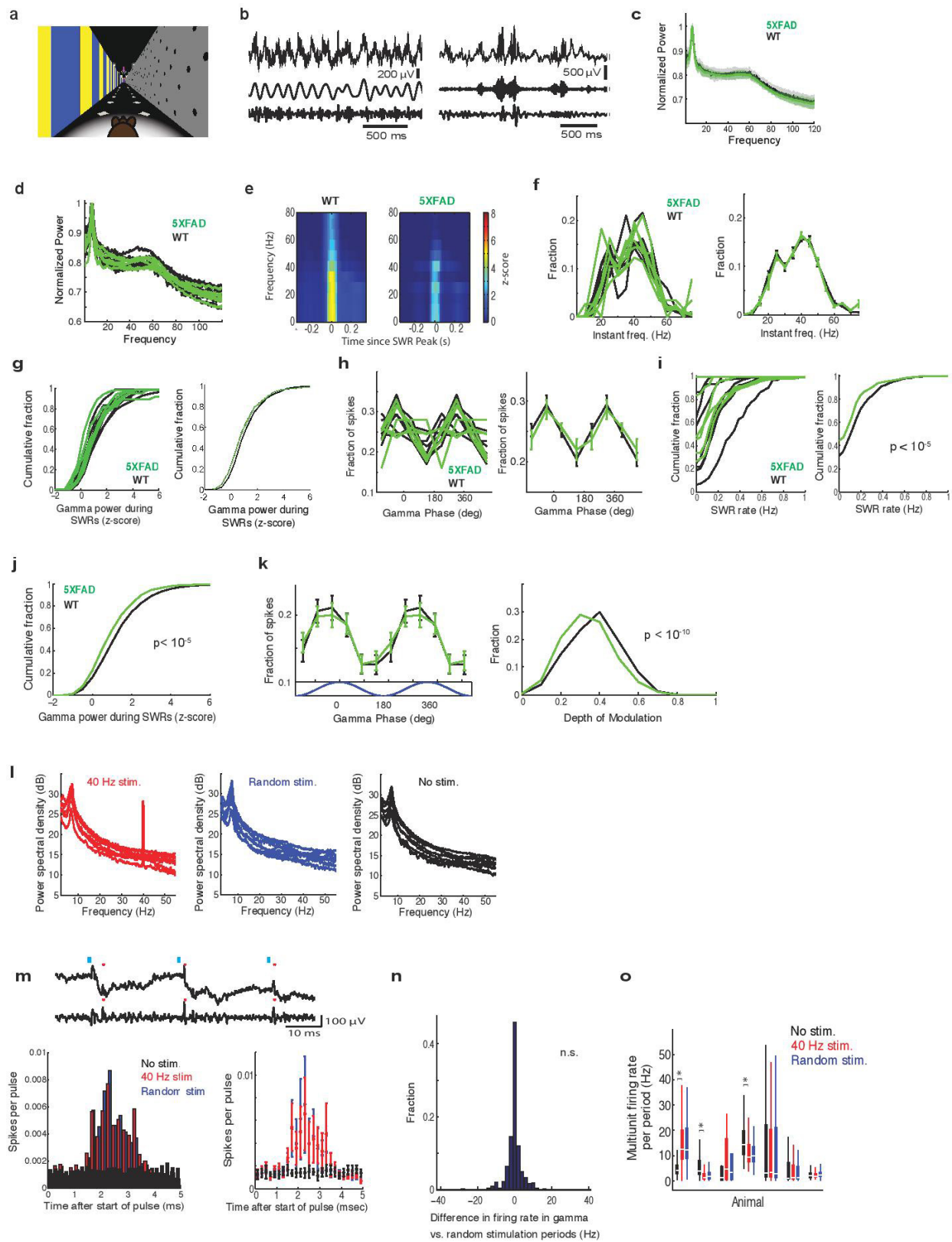
Code availability. Code is publicly available upon request from the corresponding author.

Statistics. For electrophysiological data that were not normally distributed, results are presented as medians and quartiles unless otherwise noted. Two-sided Wilcoxon rank sum tests for equal medians were performed to determine whether distributions were significantly different, and Wilcoxon signed rank tests were performed to determine whether distributions were significantly different from zero as these do not assume data are normally distributed. Variability was similar between the groups that were statistically compared. The Bonferroni method was used to correct for multiple comparisons. No statistical method was used to estimate sample size, but it was consistent with previous publications.

Molecular and biochemical results are presented as mean + s.e.m. Percentages stated are group means. All statistical analysis used Prism GraphPad software. Normality was determined using the D'Agostino and Pearson omnibus normality test. Variability was similar between the groups that were statistically compared. Comparison data for normally distributed data consisting of two groups were analysed by two-tailed unpaired *t*-tests. Comparison of normally distributed data consisting of three or more groups was by one-way ANOVA followed by Tukey's multiple comparisons test. Comparisons for non-normally distributed data were performed using Mann–Whitney tests. The statistical test, exact *P* values, and sample size (*n*) for each experiment are specified in the figure legend. For optogenetic ELISA data, two-sided unpaired Student's *t*-tests were performed to compare mice from the same litter that received different conditions. No statistical method was used to estimate sample size, but it is consistent with previous publications. Molecular and biochemical analysis used a minimum of three biological replicates per condition.

Data availability. Data are publicly available upon request from the corresponding author.

30. Harvey, C. D., Collman, F., Dombeck, D. A. & Tank, D. W. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* **461**, 941–946 (2009).
31. Aronov, D. & Tank, D. W. Engagement of neural circuits underlying 2D spatial navigation in a rodent virtual reality system. *Neuron* **84**, 442–456 (2014).
32. Rossant, C., et al. Spike sorting for large, dense electrode arrays. *Nature Neurosci.* **19**, 634–642 (2016).
33. Buzsáki, G. Theta oscillations in the hippocampus. *Neuron* **33**, 325–340 (2002).
34. Ravassard, P. et al. Multisensory control of hippocampal spatiotemporal selectivity. *Science* **340**, 1342–1346 (2013).
35. Zhang, Y. et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34**, 11929–11947 (2014).
36. Crotti, A. et al. Mutant Huntingtin promotes autonomous microglia activation via myeloid lineage-determining factors. *Nature Neurosci.* **17**, 513–521 (2014).
37. Erny, D. et al. Host microbiota constantly control maturation and function of microglia in the CNS. *Nature Neurosci.* **18**, 965–977 (2015).
38. Chiu, I. M. et al. A neurodegeneration-specific gene-expression signature of acutely isolated microglia from an amyotrophic lateral sclerosis mouse model. *Cell Reports* **4**, 385–401 (2013).
39. Gosselin, D. et al. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* **159**, 1327–1340 (2014).
40. Huang, S. C.-C. et al. Cell-intrinsic lysosomal lipolysis is essential for alternative activation of macrophages. *Nature Immunol.* **15**, 846–855 (2014).
41. Yu, H. et al. Tet3 regulates synaptic transmission and homeostatic plasticity via DNA oxidation and repair. *Nature Neurosci.* **18**, 836–843 (2015).

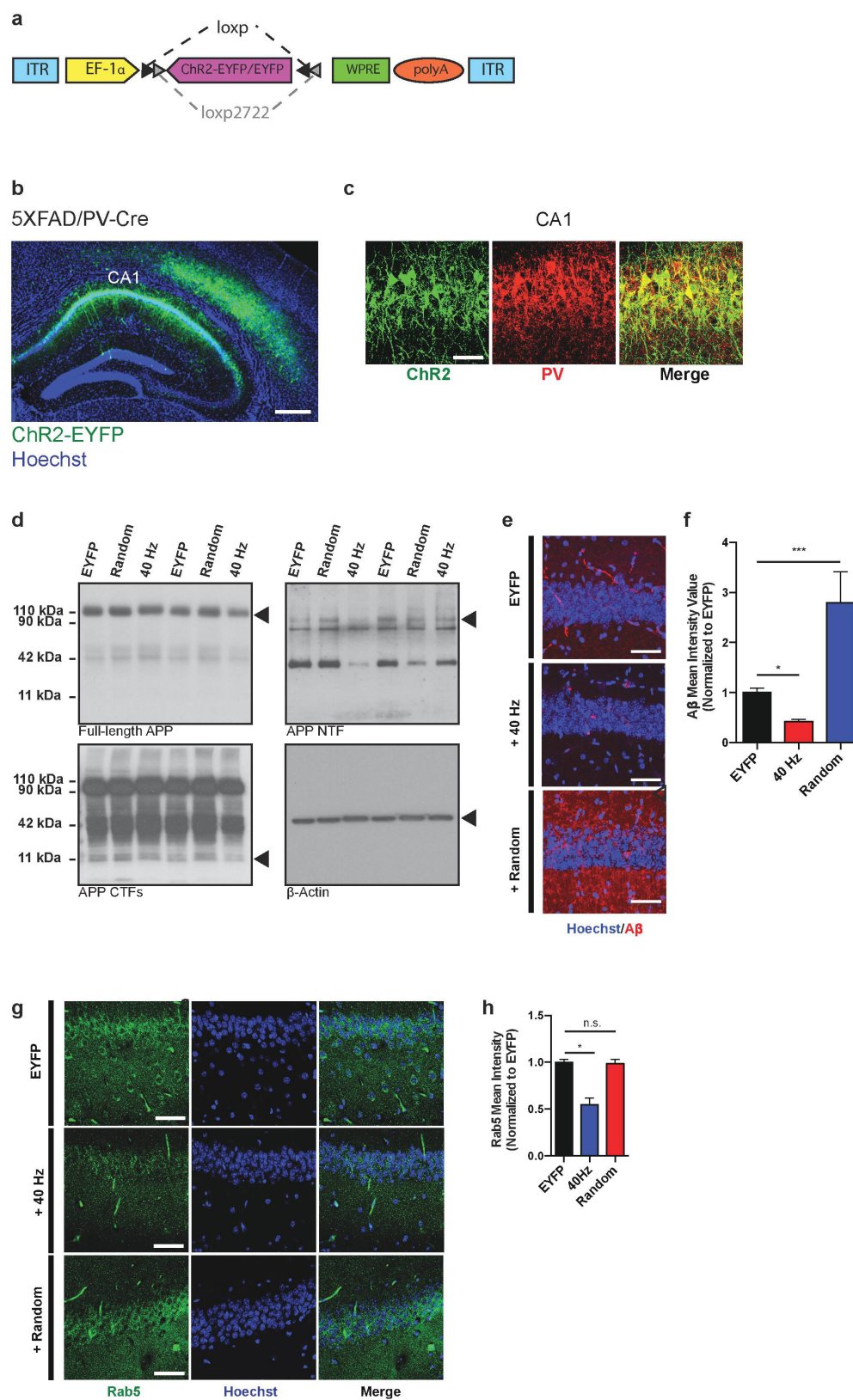


Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | 5XFAD mice have reduced power in gamma during hippocampal SWRs. **a**, Mouse in virtual reality environment.

b, LFP recorded in CA1, above, filtered for theta (left) or SWRs (right), middle, and gamma, below. **c**, Mean and standard deviation of the normalized power spectrum during theta. Each animal's power spectral density was normalized to its peak ($n = 6$ mice per group). **d**, Normalized power spectral densities during theta periods in 3-month-old 5XFAD (green, $n = 6$ mice) and WT (black/grey, $n = 6$ mice) mice. Each animal's power spectral density was normalized to its peak (in theta). **e**, Average SWR-triggered spectrograms for one WT and one 5XFAD mouse shows an increase in the gamma band during SWRs. This increase is lower in the 5XFAD mouse than in the WT mouse ($n = 370$ and 514 SWRs in WT and 5XFAD, respectively; WT mouse shown here is the same as in Fig. 1a). This range of frequencies is often called 'slow gamma' to distinguish it from faster oscillations (65–140 Hz) that have also been included in the gamma range but for which the origins are less well understood. **f**, Distributions for each recording (left) and the mean and standard error across sessions (right) of instantaneous gamma frequencies during SWRs in 5XFAD (green) and WT (black) mice show distributions around 40 Hz ($n = 820, 800, 679, 38, 1,875, 57$ gamma cycles per session in six 5XFAD animals in six recording sessions and $181, 1,075, 919, 1,622, 51, 1,860, 1,903$ gamma cycles per session in six WT animals in seven recording sessions). **g**, Cumulative distribution of the z-scored gamma power during the 100 ms around the peak of the SWR for WT (black) and 5XFAD animals (green) for each animal (left) and the mean and standard error (shaded) across animals (right) ($n = 514, 358, 430, 22, 805, 37$ SWRs per session in six 5XFAD animals and $82, 311, 370, 776, 18, 710, 818$ SWRs per session in six WT animals). **h**, Fraction of spikes in CA1 during SWRs as a function of the phase of gamma in 5XFAD (green) and WT (black) mice for each animal (left) and the mean and standard error across animals (right, $n = 2475, 1060, 3092, 25, 6521, 123$ spikes during SWRs per session in six 5XFAD mice and $360, 4741, 1,564, 2,961, 88, 3,058, 4,270$ spikes during SWRs per session in six WT mice). **i**, SWR rate per non-theta period in 5XFAD (green) and WT (black) mice for each animal (left) and all animals combined (right, rank-sum test, $P < 10^{-10}$, $n = 117, 210, 151, 55, 100, 1$ non-theta periods per session in six 5XFAD mice and $80, 68, 115, 95, 15, 159, 218$ non-theta periods per session in six WT mice). **j**, The cumulative distribution of gamma power during large SWRs (detection threshold greater than six standard deviations above the mean, Methods) shows significantly smaller increases in 5XFAD (green) than WT (black) mice (rank-sum test, $P < 10^{-5}$, $n = 1,000$ SWRs in six 5XFAD mice and $1,467$ SWRs in 6 WT mice). **k**, Fraction of spikes as a function of the phase of gamma during large SWRs (detection threshold greater than six standard deviations above the mean, Methods), mean \pm s.e.m. (left) and histogram of the depth of modulation of spiking (right) as a function of gamma phase in 3-month-old 5XFAD (green, $n = 6$ mice) and WT (black, $n = 6$ mice) mice (rank-sum test, bootstrap resampling

$P < 10^{-10}$, $n = 2500$ 5XFAD spike-gamma phase distributions and $3,000$ WT distributions). **l**, Power spectral density during 40 Hz stimulation (red, left), random stimulation (blue, centre), or no stimulation (black, right) of FS-PV-interneurons in CA1 for each mouse ($n = 4$ 5XFAD mice with $169, 130, 240, 73$ 40 Hz, $143, 129, 150, 72$ random, and $278, 380, 52, 215$ no stimulation periods per animal and $n = 3$ WT mice with $65, 93, 91$ 40 Hz, $64, 93, 90$ random, and $187, 276, 270$ no stimulation periods per animal). **m**, Above: example raw LFP trace (above) and the trace filtered for spikes (300–6,000 Hz, below), with spikes indicated with red stars after optogenetic stimulation (blue vertical lines). Below: histogram of spikes per pulse after the onset of the 1 ms laser pulse during 40 Hz stimulation (red), random stimulation (blue), or no stimulation (black, $n = 345,762$ 40 Hz, $301,559$ random pulses, and $32,350$ randomly selected no stimulation times at least 500 ms apart from 552 40 Hz, 543 random, and 1681 no stimulation periods in four 5XFAD and three WT mice). **n**, Histogram of the difference in firing rates between 40 Hz stimulation and random stimulation periods shows that both types of stimulation elicit similar amounts of spiking activity (Wilcoxon signed rank test for zero median, $P > 0.6$, $n = 538$ stimulation periods from four 5XFAD and three WT mice, NS, not significant). **o**, Multiunit firing rates per 40 Hz stimulation (red), random stimulation (blue), and no stimulation (black) period for each animal. Box-and-whisker plots show median (white lines in box) and quartiles (top and bottom of box). In all animals firing rates between 40 Hz and random stimulation were not significantly different, showing that the random stimulation condition serves as a control for spiking activity (rank-sum tests for each animal, three WT and four 5XFAD mice, $P > 0.09$, $n = 87, 130, 8, 65, 93, 91, 73$ 40 Hz stimulation periods and $85, 129, 5, 64, 93, 90, 72$ random stimulation periods per animal). We also examined whether 40 Hz stimulation caused neuronal hyperactivity relative to no stimulation, because, according to a recent report, this could have negative effects on neural circuit function²⁶. In most animals the firing rates between 40 Hz or random stimulation and no stimulation were not significantly different (rank-sum tests for each animal, two WT and two 5XFAD, $P > 0.25$, $n = 8, 93, 91, 73$ 40 Hz stimulation periods and $15, 277, 270, 215$ baseline periods per animal) or the firing rates during 40 Hz or random stimulation were lower than during no stimulation (rank-sum tests for each animal, one WT and one 5XFAD, $P < 10^{-5}$, which is significant when corrected for performing multiple comparisons, $n = 130, 65$ 40 Hz stimulation periods and $379, 187$ baseline periods per animal) indicating that 40 Hz stimulation did not cause neuronal hyperactivity. In one animal there was significantly more activity with 40 Hz or random stimulation than during baseline (rank-sum test for one 5XFAD, mouse, $P < 10^{-5}$, $n = 87$ 40 Hz stimulation periods and 251 baseline periods per animal). Therefore in six out of seven animals we see no evidence that the 40 Hz optogenetic stimulation of FS-PV-interneurons causes hyperactivity.



Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | ChR2 was expressed in FS-PV-interneurons for optogenetic stimulation. **a**, AAV-DIO-ChR2-eYFP or AAV-DIO-eYFP drives Cre-dependent expression of ChR2-eYFP or eYFP to produce cell-type-specific targeting of ChR2 or eYFP, respectively. In the presence of Cre, ChR2-eYFP or eYFP is inverted into the sense direction and expressed from the EF-1 α promoter in PV-positive cells. ITR, inverted terminal repeat; polyA; WPRE, woodchuck hepatitis B virus post-transcriptional element. **b**, ChR2-eYFP was strongly expressed in PV-positive interneurons in CA1 of 3-month-old 5XFAD/PV-Cre mice (scale bar, 100 μ m). **c**, Immunohistochemistry with anti-eYFP and anti-PV antibodies in CA1 of 3-month-old 5XFAD/PV-Cre mice expressing AAV-DIO-ChR2-eYFP shows eYFP expression only in PV-positive cells (scale bar, 50 μ m). **d**, Representative western blots showing levels

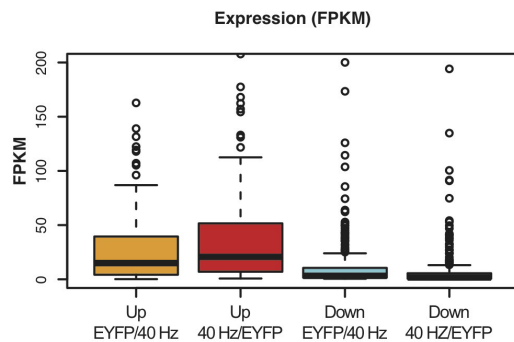
of full-length APP (top left, CT695), APP CTFs (bottom left, CT695), APP NTF (top right, A8967), and β -actin (bottom right, A5316, loading control) in CA1 in eYFP, random, and 40 Hz stimulation conditions, one mouse per lane, with two biological replicates of each condition. **e**, Immunohistochemistry with anti-A β (12F4, red) antibodies in CA1 of 5XFAD/PV-Cre mice expressing only eYFP or ChR2 with 40 Hz, and random stimulation conditions (scale bar, 50 μ m). **f**, Bar graphs represent the relative immunoreactivity of A β normalized to eYFP ($n = 4$ mice per group; $*P < 0.05$, $***P < 0.001$ by one-way ANOVA). Bar graphs show mean \pm s.e.m. **g**, Immunohistochemistry with anti-Rab5 antibody (ADI-KAP-GP006-E, green) in CA1 of 5XFAD/PV-Cre mice (scale bar, 50 μ m). **h**, Relative Rab5 intensity levels normalized to eYFP controls ($n = 3$ mice per group).

a Up-regulated genes

2010002N04Rik	Beta-s	Card10	Cyp2d22	Gldc	H1fx	Igfbp4	Lag3	Mpped1	Pcsk1n	Prrt1	Slc12a9	Tmem198
2010300C02Rik	Bst2	Cd68	Dcackd	Gm129	H2-D1	Igfbp1	Lcat	Mt1	Pdzd2	Rab40b	Slc25a34	Tpst2
2410018L13Rik	C1qa	Cebpb	Egr4	Gm2115	H2-K1	Irf7	Lefty1	Mt2	Pqls	Rara	Slc29a4	Trim30a
Adra2c	C1qb	Cebpd	Erf	Gng7	Hipk4	Irf9	Lgals3bp	Mtap1s	Phyhd1	Rasl11b	Spp1	Ttr
Aqf2	C1qc	Cirbp	F730043M19Rik	Gpnmb	Hmox1	Itpka	Lingo3	Npy	Pitpnm2	Rbm3	Spsb1	Unc5a
Aqxt2l1	C1ql2	Cnn2	Fam107a	Gpr25	I830012O16Rik	Junb	Lrq1	Nr1d1	Plekhg5	Rpph1	Ssbp4	Uqcr11
Arc	C1qtnf4	Colt1	Fam163b	Gpr37l1	Icam1	Kcnc4	Ltbp4	Nr4a1	Pnpla7	Rprml	Sstr4	Usp18
Atf3	C3ar1	Crip2	Fmo2	Grm2	Icam5	Kcnh3	Lyz2	Oasl2	Pou3f1	Sbk1	Tfcp2l1	Vwf
B2m	C4b	Cst3	Fn1	Gstm1	Ifit1	Kcnj4	Metm	Palm	Ppp1r1a	Scara3	Thbs4	Wfs1
BC018242	Car7	Ctnx1	Gbp3	Gstm6	Ifit3	Klf16	Mmp12	Parp14	Prr7	Sh3bgrl3	Thrsp	Xdh

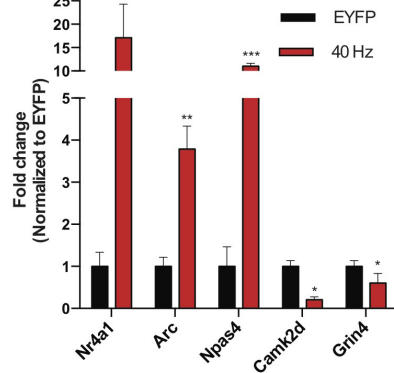
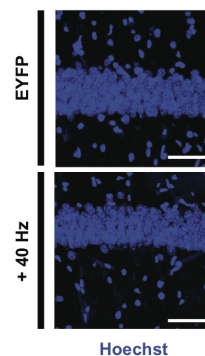
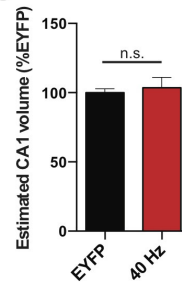
b Down-regulated genes

1700003M02Rik	A630089N07Rik	Arhgap24	Ccdc141	Cited2	Edi3	Fam19a4	Gpr115	Hs6st2	Kcnma1	Lrrc48	Necab2	Pcsk1	Prkcd	Rgs16	Sh3bgrl2	Spat18	Th	Tyrrp1	Ysk4
1700007K13Rik	AF529169	Asb2	Ccdc153	Clec18a	Efcab1	Fbln7	Gpr123	Hsp90aa1	Kcnk3	Lrrc55	Necab3	Pdp1	Prkch	Rgs3	Shox2	Spock1	Timp2	Ubxn10	Zcchc12
1700009P17Rik	AU021034	Aspa	Ccdc19	Clic6	Efnf5	Faf1	Gpr139	Hsp90b1	Kcnk10t1	Malat1	Nexn	Peg10	Prkca	Rgs4	Shroom3	Spock3	Tm4sf1	Uat8a	Zdbf2
1700026D08Rik	AW551984	Balap3	Ccdc3	Cntn6	Eif5a2	Fgf10	Gpr151	Hspa4l	Kctd12b	Mcf2	Nfam1	Pgap1	Prkq2	Rgs6	Slc12a2	Srgap1	Tmem130	Unc13c	Zdhhc22
1700027A23Rik	Adamts15	Bbox1	Ccdc40	Cntnap4	Elavl2	Fhd1	Gpr153	Htr2c	Kctd8	Medf11	Nab	Pabp1	Prokr2	Rims3	Slc17a6	Sltsia2	Tmem132c	Vangl1	Zfhx3
1700028P14Rik	Adamts9	Bmp7	Ccdc81	Cobl	Elavl4	Foxi1	Gpr26	Htr5b	Klf9	Mgat4c	Nhlh2	Phactr2	Prr5l	Rit2	Slc38a1	Strbp	Tmem163	Vat1	Zfp474
1700040L02Rik	Adcyap1	Btdb11	Cd109	Coch	Eifn1	Foxp2	Gpr4	Htr7	Klt	Mif1	Nopa	Pirt	Prrq4	Rnf152	Slc39a4	Sv2b	Tmem176a	Vat1l	Zfp618
1700094D03Rik	Adra1b	C53008M17Rik	Cd24a	Col12a1	Emb	Frem3	Gprasp2	Hvdl	Klil	Mme	Nor1	Pkib	Ptqds	Robo1	Slc5a3	Synpo2	Tmem212	Vav2	Zfp941
1810041L15Rik	Acabp1	Cacna2d2	Cdh26	Col8a2	Enkur	Fstl5	Gpx3	Inadl	Klhl1	Mob3b	Nr4a2	Plagl1	Pttn14	Rorb	Slc5a7	Svt15	Tmem56	Vav3	Zic1
2310039L15Rik	Aqt	Calb2	Cdhr3	Cpne4	Eno4	Fzd1	Grb10	Igca	Lars2	Mreg	Nrip3	Pcb4	Pvrl2	Rps6ka8	Slc6a9	Svt4	Tnc	Vwa5b1	Zic2
2410004P03Rik	Aif1l	Calr3	Cdr1	Cpne9	Enox2	Fzd10	Gria4	Igub	Lbh	Mrv1	Nrp2	Pch1	Pvrl3	Rsp1	Slc7a11	Svt6	Tnnt1	Wbscr27	Zic3
3110047P20Rik	Ak4	Camk2d	Cdr2	Dach1	Epha8	Galnt13	Grid2ip	Irx1	Lbp	Msi2	Nrsn2	Pch2	Rab37	Rsp4a	Slc9a4	Svt9	Trhde	Wdr16	Zic4
3632451O06Rik	Ak7	Car10	Chat	Dcn	Epn3	Gap43	Gria3	Irx2	Ldhd	Mtrf1	Ntnq1	Picx2	Ramp3	Rspo2	Slco2a1	Tac1	Trim36	Wdr52	Zic5
4930451C15Rik	Akap12	Cast	Chgb	Ddc	Ernn	Galm	Grk4	Irx3	Lect1	Mum11l	Nudt4	Pld5	Ranbp3l	Scn1a	Sllt2	Tac2	Trim66	Wdr6	
4932411L15	Amigo2	Chn1	Chrd1	Dnahc5	Etv1	Gdpd5	Grm4	Iqa3	Lef1	Musk	Olfm3	Plekhg1	Rap1gap	Scube1	Slltrk6	Tacr1	Trpc3	Wdr78	
4932425I24Rik	Amotl1	Cbln2	Chrna3	Dnahc6	Eph5	Gfra1	Gucy1a3	Kcnc2	Lhfp1	Myb	Otn	Ptxnc1	Rasgef1b	Scube3	Sncg	Tcf7l2	Trps1	Wdr96	
5730508B09Rik	Ankrd29	Cbln4	Chrna4	Dpp10	Fabp7	Gm6300	Hcn4	Kcnc4	Lhfp3	Mycbpap	Qtx2	Popdc3	Rassf9	Sema3d	Sntn	Tekt1	Tsku	Wfikk2	
6330406I15Rik	Ankrd34c	Chmb3	Dpy19l1	Fam149a	Gm7609	Hdc	Kenip1	Lhx9	Myoc	Pamr1	Pou4f1	Rbms3	Sema6a	Socs2	Tex15	Ttc18	Wf1		
A2m	Ano1	Ccdc135	Chmb4	Dynlrb2	Fam196b	Gm973	Hhip	Kcnj12	Lrauk	Ndn	Pbx3	Ppp1r32	Resp18	Serpinf1	Sox5	Tex9	Ttc21a	Wls	
A330021E22Rik	Aqp4	Ccdc136	Cit	Ebf1	Fam198b	Gng8	Hivep1	Kcnj16	Lrrc23	Necab1	Pcp4	Ppp1r36	Ret	Sgpp2	Spag16	Tgfb2	Ttc39a	Wnt3	

c**d****Enrichment of up-regulated genes under different perturbations**

Name of perturbed transcriptome	NES	Normal p-value	FDR q-value
MCSF treated microglia	1.76	0.000	0.000
NMDA treated neurons	1.62	0.000	0.000
IL34 treated microglia	1.59	0.000	0.000
GMCSF treated microglia	1.49	0.005	0.004
Bicuculline treated neurons	1.49	0.016	0.013
ALS SOD1 mutant microglia	-1.26	0.050	0.028
LPS&IFNγ treated macrophage (M1)	1.18	0.122	0.081
MeCP2 null microglia	1.16	0.164	0.127
IL4 treated macrophage (M2)	-1.19	0.101	0.147
Huntington HTT mutant microglia	1.03	0.371	0.361
Germ-free microglia	0.94	0.604	0.667
Tetradotoxin treated neurons	-0.76	0.941	0.970

NES: normalized enrichment score

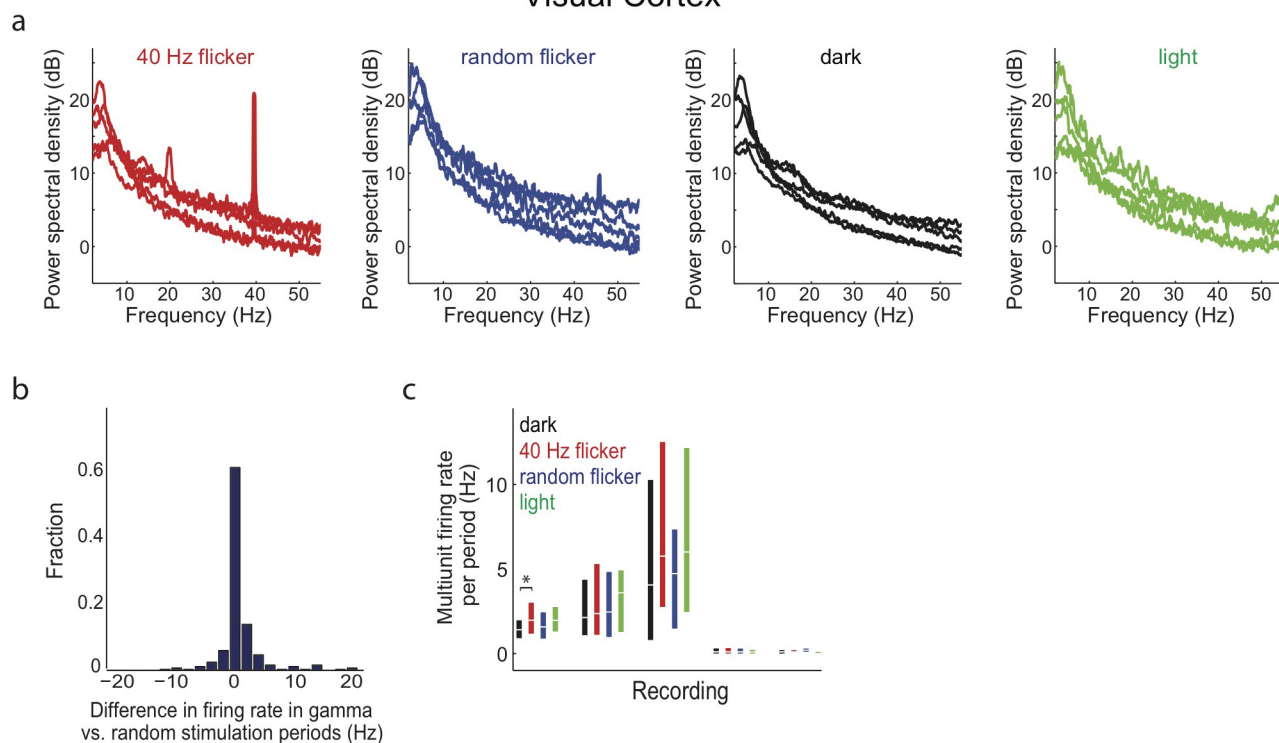
e**f****g**

Extended Data Figure 3 | See next page for caption.

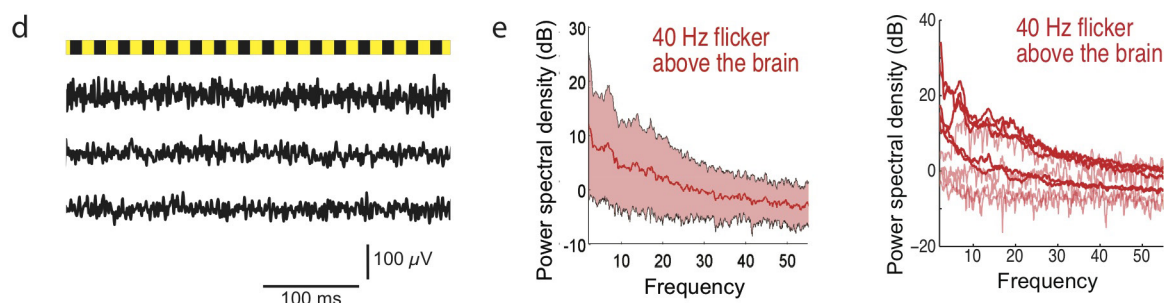
Extended Data Figure 3 | Optogenetically driven 40 Hz oscillations in CA1 cause changes in gene regulation and immediate early gene expression. **a**, Table of 130 genes upregulated by 40 Hz FS-PV-interneuron stimulation determined by whole-transcriptome RNA-seq of CA1 from 3-month-old 5XFAD/PV-Cre mice ($P < 0.05$ by Cufflinks 2.2). **b**, Table of 393 genes downregulated by 40 Hz FS-PV-interneuron stimulation determined by whole transcriptome RNA-seq of CA1 from 3-month-old 5XFAD/PV-Cre mice ($P < 0.05$ by Cufflinks 2.2). **c**, Box plot showing FPKM values of up- and downregulated genes in eYFP and 40 Hz groups. Box shows median (black lines in box) and quartiles (top and bottom of box), whiskers show minimum and maximum values, and circles show outliers. **d**, GSEA statistics tables showing statistical significance of correlation between genes up- or downregulated by 40 Hz stimulation and publicly available neuron-, microglia-, and macrophage-specific RNA-seq

data under different chemical and genetic perturbations; the perturbation terms were ranked on the basis of the false discovery rate q values for the upregulated gene list, from the smallest to the largest (Methods). **e**, RT-qPCR verification of specific gene targets in the RNA-seq dataset. Bar graph shows relative RNA levels (fold change) from eYFP (black) and 40 Hz stimulation (red) conditions ($*P < 0.05$, $**P < 0.01$, $***P < 0.001$ by Student's t -test, $n = 3$ mice per group). All bar graphs show mean + s.e.m. **f**, Immunohistochemistry with Hoechst to label all cell nuclei in CA1 of 5XFAD/PV-Cre mice expressing only eYFP or ChR2 with 40 Hz stimulation conditions (scale bar, 50 μm). **g**, Bar graph represents the estimated CA1 thickness in 5XFAD/PV-Cre mice expressing only eYFP or ChR2 with 40 Hz stimulation conditions ($n = 4$ mice per group; NS, not significant by Student's t -test).

Visual Cortex



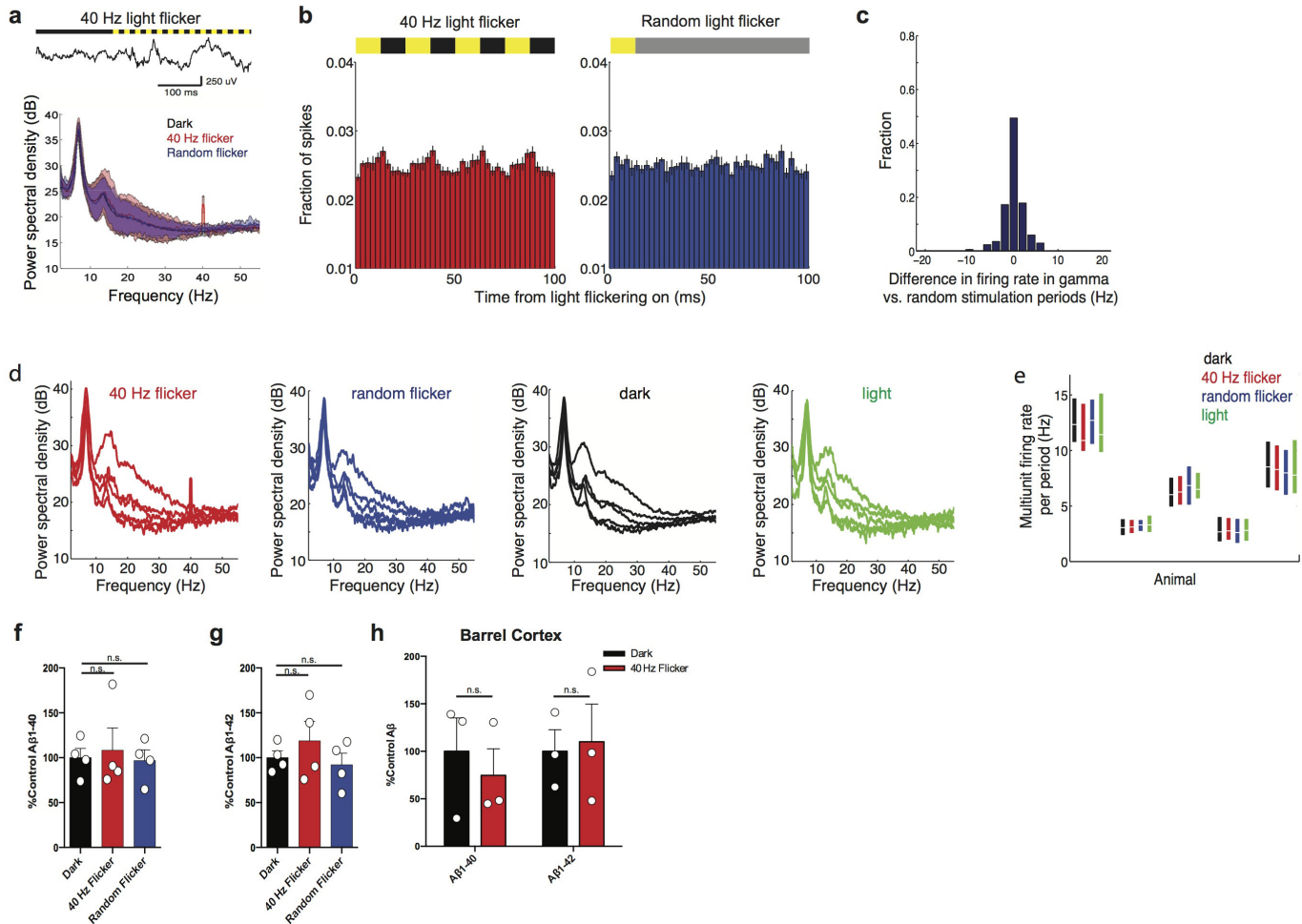
Above Brain



Extended Data Figure 4 | A 40 Hz light flicker drives 40 Hz oscillations in VC, while random flickering does not. **a**, Power spectral densities of LFPs in VC during 40 Hz light flicker (red, far left), random light flicker (blue, centre left), dark (black, centre right), or light (green, far right) in VC for each recording session for each mouse ($n = 5$ recordings from four 5XFAD mice with 47, 51, 64, 49, 16 40 Hz flicker, 47, 50, 64, 50, 16 random flicker, 279, 301, 382, 294, 93 dark and 47, 50, 64, 49, 15 light periods). Light flicker at other frequencies increased power in the flicker frequency, as others have found previously^{23,24} (data not shown). **b**, Histogram of the difference in firing rates between 40 Hz light flicker and random light flicker ($n = 226$ stimulation periods from five recording sessions in four 5XFAD mice). **c**, Multiunit firing rates in VC during 40 Hz light flicker (red), random light flicker (blue), dark (black), or light (green) periods. Box plots show median (white lines in box) and quartiles (top and bottom of box). In all animals, firing rates between 40 Hz flicker and random flicker conditions were not significantly different showing that the random stimulation condition serves as a control for spiking activity (rank-sum tests for each of five recording session from four 5XFAD mice, $P > 0.06$, $n = 47, 51, 64, 49, 16$ 40 Hz flicker periods and 47, 50, 64, 50, 16 random flicker periods per recording). There were no significant differences in firing rates between 40 Hz flicker and light conditions, indicating that

40 Hz light flicker generally did not cause neuronal hyperexcitability (rank-sum tests for each of five recording sessions from four 5XFAD mice, $P > 0.2$ for four recording sessions, $P < 0.01$ for one recording session, which is not significant when corrected for performing multiple comparisons, $n = 47, 51, 64, 49, 16$ 40 Hz periods and 47, 50, 64, 49, 16 light periods per recording). In one session, there was more activity in the 40 Hz flicker than in the dark condition. **d**, Example traces of LFPs recorded above the brain during light flicker (above, yellow indicates light on, black indicates light off), during three different recording sessions. **e**, Power spectral densities of LFPs recorded above the brain during 40 Hz light flicker show no increase in power at 40 Hz. Thus, the effect is not due to photoelectric effects on recording equipment or electrical noise ($n = 4, 2, 1, 1, 17, 42, 36, 55, 53$ 40 Hz flicker periods from four recording sessions in three 5XFAD mice undergoing VC recordings and from five recording sessions in two 5XFAD and three WT mice undergoing hippocampal recordings). Mean (solid line) and standard deviation (shaded area) across recordings are shown on the left and per animal on the right. Recordings with fewer than three flicker periods (light red) resulted in noisier power spectral densities than recordings with more data (dark red) but none showed evidence of peaks at 40 Hz.

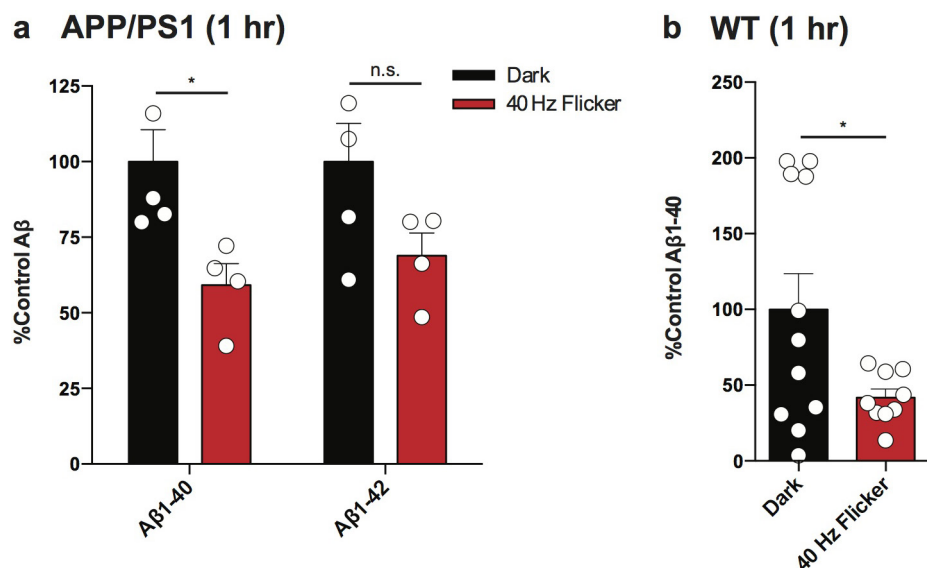
Hippocampus



Extended Data Figure 5 | A 40 Hz light flicker does not affect A β levels in hippocampus or barrel cortex.

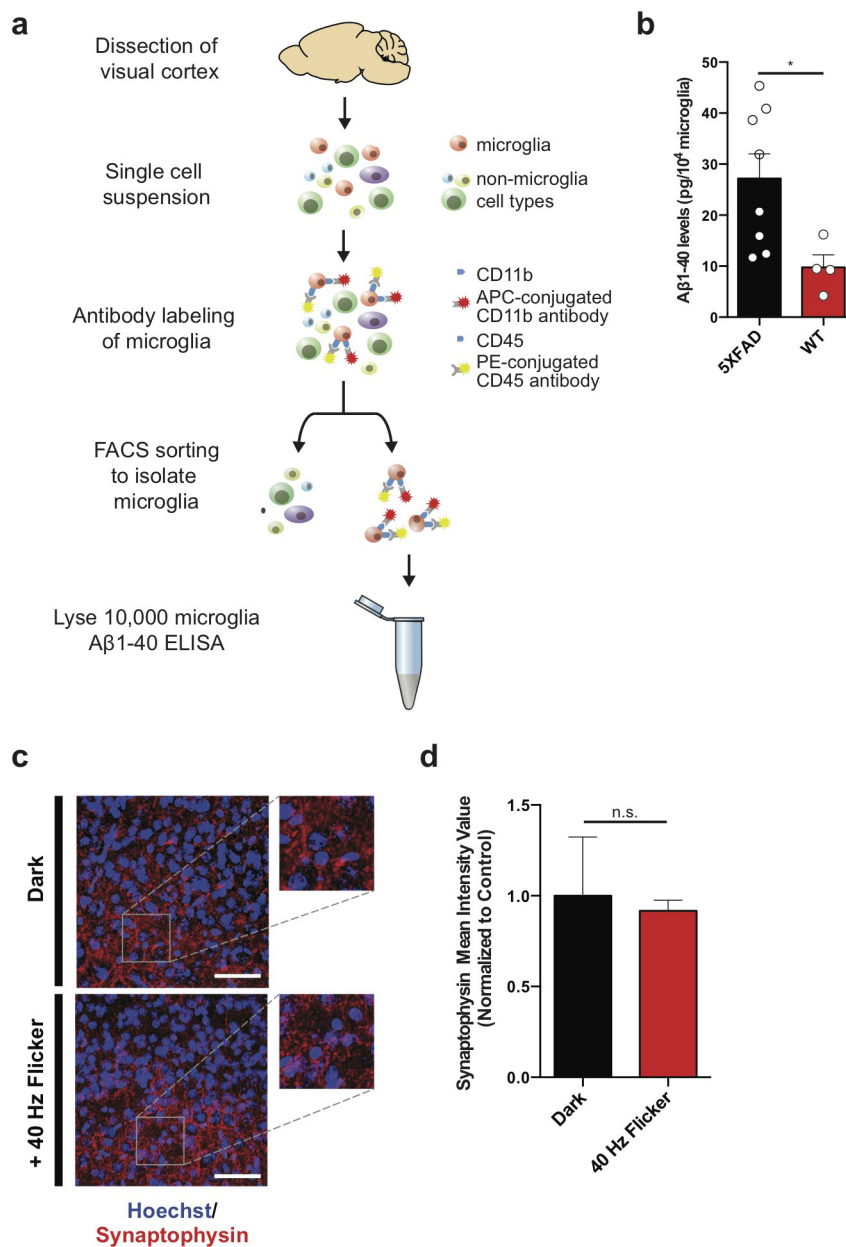
a, Example LFP trace in hippocampal CA1 before and during 40 Hz light flicker (above). Mean (solid line) and standard deviation (shaded area) of power spectral density during 40 Hz light flicker (red), random light flicker (blue), or dark (black) in CA1 ($n = 2$ 5XFAD and 3 WT mice). **b**, Histogram of the fraction of spikes in hippocampus as a function of time for four cycles of 40 Hz light flicker (left, red) or the equivalent period of time for random light flicker (right, blue, $n = 2$ 5XFAD and 3 WT mice, mean \pm s.e.m. across animals). Bar above indicates when light was on (yellow) or off (black). For random stimulation, spiking was aligned to the start of the light turning on; additional periods with light-on occurred at random intervals are indicated by grey (Methods). **c**, Histogram of the difference in firing rates between 40 Hz light flicker and random light flicker (bottom $n = 168$ stimulation periods from five recording sessions in two 5XFAD and three WT mice). **d**, Power spectral densities of LFPs in CA1 during 40 Hz light flicker (red, far left), random light flicker (blue, centre left), dark (black, centre right), or light (green, far right) for each recording session for each mouse ($n = 5$ recordings from two 5XFAD and three WT mice with 22, 54, 42, 71, 55, 40 Hz flicker, 12, 34, 32, 54, 36 random flicker, 115, 240, 224, 342, 282 dark and 12, 33, 33, 54, 35 light periods). **e**, Multiunit firing rates in CA1 during 40 Hz light flicker (red), random light flicker (blue), dark (black), or light (green) periods. Box plots show median (white lines

in box) and quartiles (top and bottom of box). In all animals firing rates between 40 Hz flicker and random flicker conditions were not significantly different, showing that the random stimulation condition serves as a control for spiking activity (rank-sum tests for each of five recordings from two 5XFAD and three WT animals, $P > 0.2$, $n = 22, 54, 42, 71, 55$ 40 Hz flicker periods and 12, 34, 32, 54, 36 random flicker periods per recording). There were no significant differences in firing rates between 40 Hz flicker and light conditions indicating that 40 Hz light flicker generally did not cause neuronal hyperexcitability (rank-sum tests for each of five recordings from two 5XFAD and three WT animals, $P > 0.3$, $n = 22, 54, 42, 71, 55$ 40 Hz periods and 12, 34, 33, 54, 35 light periods per recording). **f**, Bar graphs of relative $A\beta_{1-40}$ levels in VC of 5XFAD mice in dark, 40 Hz flicker, and random flicker conditions, normalized to dark ($n = 4$ mice per group; NS, not significant). Bars represent mean \pm s.e.m. Circles superimposed on bars in bar graphs indicate individual data points in each group. **g**, Bar graphs of relative $A\beta_{1-42}$ levels in VC of 5XFAD mice in dark, 40 Hz flicker, and random flicker conditions, normalized to dark ($n = 4$ mice per group; NS indicates not significant). Bars represent mean \pm s.e.m. Circles superimposed on bars indicate individual data points in each group. **h**, Bar graph of relative $A\beta_{1-40}$ and $A\beta_{1-42}$ levels in barrel cortex of 5XFAD mice in dark and 40 Hz flicker conditions, normalized to dark ($n = 3$ mice per group; NS, not significant by Student's t -test).



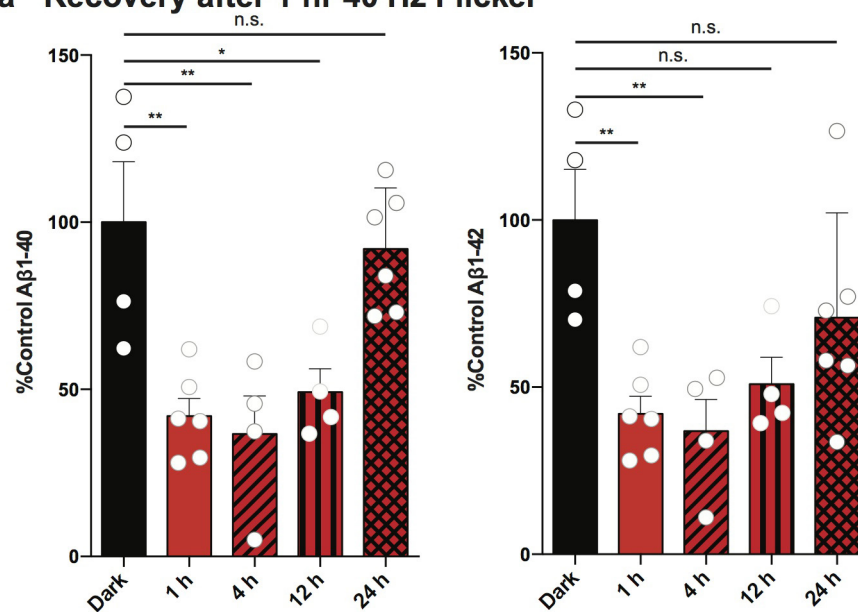
Extended Data Figure 6 | Acute reduction in Aβ after light flicker in APP/PS1 and WT mice. **a**, Bar graph of relative Aβ₁₋₄₀ and Aβ₁₋₄₂ levels of APP/PS1 in VC in dark and 40 Hz flicker conditions, normalized to dark ($n = 4$ mice per group for dark and $n = 4$ mice per group for 40 Hz flicker conditions; NS, not significant, $*P < 0.05$, by Student's *t*-test). All bar graphs show mean + s.e.m. throughout this figure. Circles superimposed

on bars in bar graphs indicate individual data points in each group. **b**, Bar graph of relative mouse Aβ₁₋₄₀ and Aβ₁₋₄₂ levels in VC of 9-month-old WT mice in dark and 40 Hz flicker conditions, normalized to dark ($n = 11$ mice per group for dark and $n = 9$ mice per group for 40 Hz flicker conditions; $*P < 0.05$, by Student's *t*-test).

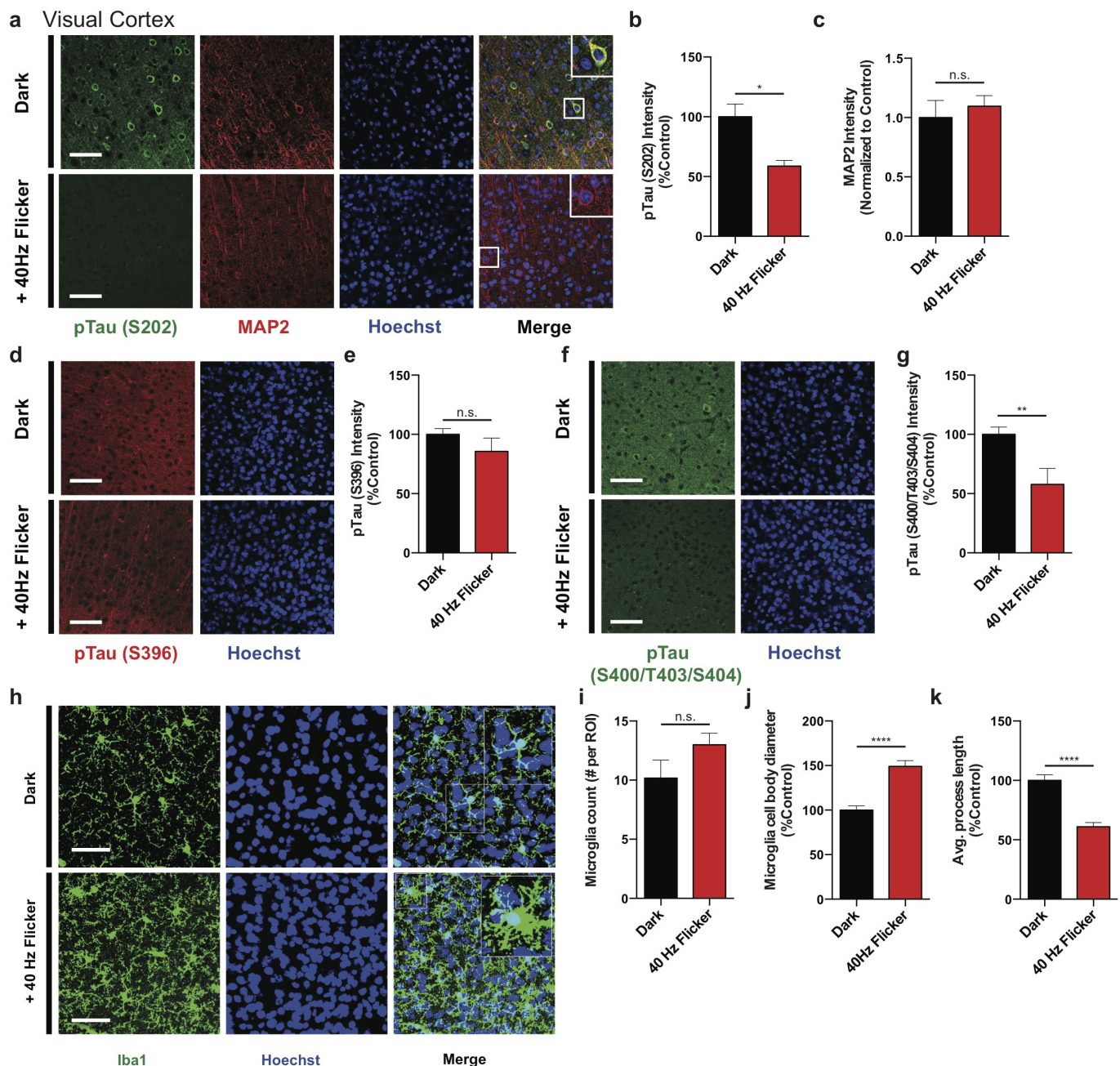


Extended Data Figure 7 | A 40 Hz light flicker does not decrease synaptic density in VC. **a**, Schematic depicting isolation of microglia from VC. VC was dissected, then single cells were suspended and labelled with CD11b and CD45 antibodies. Subsequently, cells were sorted by FACS and lysed. A β 1-40 levels were analysed by ELISA. **b**, Bar graph of A β 1-40 levels in microglia purified using FACS (Methods) from VC of 3-month-old 5XFAD and WT mice ($n = 8$ mice per group for 5XFAD and $n = 4$ mice per group for WT mice; $*P < 0.05$ by Student's t -test).

Circles superimposed on bars indicate individual data points in each group. **c**, Immunohistochemistry with SVP38 (red) antibodies to detect synaptophysin in VC of 3-month-old 5XFAD mice in dark and 40 Hz flicker conditions (images were taken with 40 \times objective; scale bar, 50 μ m). Right: 100 \times rendering of dark and 40 Hz flicker conditions. **d**, Bar graph of relative SVP38 intensity levels in VC of 5XFAD mice after dark (black) and 40 Hz (red) flicker conditions, normalized to dark ($n = 4$ mice per group; NS, not significant by Student's t -test).

a Recovery after 1 hr 40 Hz Flicker

Extended Data Figure 8 | Aβ levels in VC return to baseline 24 h after 1 h of 40 Hz light flicker. a, Bar graph of relative Aβ₁₋₄₀ and Aβ₁₋₄₂ levels in VC of 5XFAD mice 1, 4, 12, and 24 h after 1 h of dark or 40 Hz flicker treatment, normalized to dark ($n = 4$ mice per group for 4 and 12 h wait, $n = 6$ for 1 and 24 h wait, $n = 12$ for dark; NS, not significant, $*P < 0.05$, $**P < 0.01$ by one-way ANOVA).



Extended Data Figure 9 | Driving 40 Hz oscillations in VC via light flicker reduces phosphorylated tau in a tauopathy mouse model.

a, Immunohistochemistry with anti-pTau (S202, green) and anti-MAP2 (red) antibodies in VC of 4-month-old P301S mice after 7 days of 1 h per day dark or 40 Hz flicker conditions (images were taken with 40× objective; scale bar, 50 μm). Insets: 100× rendering of representative cell body in dark and 40 Hz flicker conditions. No changes were observed by western blot (data not shown). **b**, Bar graph of relative pTau (S202) intensity levels in P301S mice after 7 days of 1 h per day dark (black) and 40 Hz flicker (red) conditions ($n = 8$ mice per group; $*P < 0.05$ by Student's *t*-test). All bar graphs show mean + s.e.m. throughout. **c**, Bar graph of relative MAP2 intensity levels in P301S mice after 7 days of 1 h per day dark (black) and 40 Hz flicker (red) conditions ($n = 8$ mice per group; NS, not significant by Student's *t*-test). **d**, Immunohistochemistry with anti-pTau (S396, red) antibodies in P301S mice after 7 days of 1 h per day dark and 40 Hz flicker conditions (scale bar, 50 μm). **e**, Bar graph of relative pTau (S396) fluorescence intensity levels in P301S mice after 7 days of 1 h per day dark (black) and 40 Hz flicker (red) conditions ($n = 8$ mice per group; $****P < 0.0001$ by Student's *t*-test). **f**, Immunohistochemistry with

anti-pTau (S404, green) antibodies in P301S mice after 7 days of 1 h per day dark and 40 Hz flicker conditions (scale bar, 50 μm). **g**, Bar graph of relative pTau (S400/T403/S404) fluorescence intensity levels in P301S mice after 7 days of 1 h per day dark (black) and 40 Hz flicker (red) conditions ($n = 8$ mice per group; $**P < 0.01$ by Student's *t*-test). Bar graphs show mean + s.e.m. **h**, Immunohistochemistry with anti-Iba1 (019-19741, green) antibodies in 4-month-old P301S mice after 7 days of 1 h per day dark and 40 Hz flicker conditions (images were taken with 40× objective; scale bar, 50 μm). Insets: 100× rendering of representative microglia in eYFP and 40 Hz stimulation conditions. **i**, Bar graph of the number of microglia in P301S mice after 7 days of 1 h per day dark (black) and 40 Hz flicker (red) conditions ($n = 8$ mice per group; NS, not significant by Student's *t*-test). **j**, Bar graph of the diameter of microglia cell bodies in P301S mice normalized to dark after 7 days of 1 h per day dark (black) and 40 Hz flicker (red) conditions ($n = 8$ mice per group; $****P < 0.0001$ by Student's *t*-test). **k**, Bar graph of the average length of microglia primary processes in P301S mice normalized to control after 7 days of 1 h per day dark (black) and 40 Hz flicker (red) conditions ($n = 8$ mice per group; $****P < 0.0001$ by Student's *t*-test).

Extended Data Table 1 | Raw A β_{1-40} and A β_{1-42} concentrations

Treatment	Dilution Factor	Average A β_{1-40} Concentration (pg/ml)	Average A β_{1-42} Concentration (pg/ml)
Optogenetics			
PV-Cre EYFP	1:2	100.01, 61.598, 65.462, 82.509, 69.023, 70.831, 82.152, 74.314	58.777, 54.546, 30.585
PV-Cre 40 Hz	1:2	46.604, 31.041, 26.639, 55.612, 69.326, 17.711, 3.9951	27.271, 41.950, 18.790, 18.262
PV-Cre 8 Hz	1:2	101.268, 54.283, 90.190, 151.690	50.699, 122.85, 35.507
PV-Cre Random	1:2	235.68, 89.962, 157.37, 323.902, 451.78, 241.63	54.029, 137.78, 144.63
α CaMKII-Cre EYFP	1:2	45.813, 59.069, 40.404, 66.810	72.052, 36.573, 67.243, 59.295
α CaMKII-Cre 40 Hz	1:2	55.942, 44.270, 57.498, 47.382, 115.08, 75.673	70.847, 79.683, 61.429
α CaMKII-Cre 8 Hz	1:2	52.829, 46.604, 57.720	95.939, 21.640, 102.987
α CaMKII-Cre Random	1:2	218.00, 191.72, 159.07	66.203, 168.867, 176.404
Light flicker			
Dark 1 hr VC	1:2	343.8, 245.3, 210.6, 343.8, 588.4, 394.9, 123.3, 336.3, 328.2, 579.1, 420.0, 339.2	449.5, 320.7, 275.2, 449.5, 769.2, 516.2, 449.4, 320.6, 275.2, 449.4, 769.1, 516.1
Light 1 hr VC	1:2	366.9, 632.4, 378.2, 314.1, 266.9, 264.1	616.4, 592.3, 802.9, 394.5, 330.7, 337.8
20 Hz 1 hr VC	1:2	944.4, 313.2, 595.9, 530.0, 456.5, 289.9	1624, 302.4, 816.0, 687.2, 676.6, 343.0
40 Hz 1 hr VC	1:2	146.4, 143.6, 104.9, 99.6, 179.7, 219.8	191.4, 187.7, 137.2, 130.2, 234.9, 287.3
80 Hz 1 hr VC	1:2	332.5, 328.7, 363.5, 390.6, 530.0, 673.3	558.3, 418.9, 510.7, 609.5, 1186, 921.9
40 Hz + PTX 1 hr VC	1:2	367.2, 431.4, 445.2, 392.4, 386.7, 445.2	396.6, 540.5, 532.7, 705.0, 104.5, 104.5
Random 1 hr VC	1:2	461.8, 100.2, 9.819, 416.6	423.9, 157.9, 389.9, 841.5
Dark 1 hr HPC	1:2	97.949, 107.33, 119.92, 139.33	499.30, 355.13, 469.53, 598.03
40 Hz 1 hr HPC	1:2	88.136, 104.78, 161.52, 197.36	364.53, 408.41, 436.62, 873.83
Random 1 hr HPC	1:2	95.816, 136.77, 70.004, 125.47	466.39, 500.87, 311.26, 582.355
Dark 7 days soluble	1:50	1216.9, 1181.3, 1173.4, 1199.5, 134.73, 151.34, 113.26, 145.14, 127.91, 127.48, 143.02, 127.48, 141.07	5217.2, 8057.9, 9051.3, 6773.7, 244.11, 236.96, 235.38, 240.62, 286.19, 8.382, 11.21, 14.03, 13.56
Dark 7 days insoluble	1:100	1173.2, 1208.2, 1205.3, 1214.6, 994.86, 1059.2, 1176.6, 1065.4, 1002.9, 306.16, 690.70, 3442.7, 152.73	8572.7, 9127.1, 6349.3, 10138, 6852.2, 7056.7, 7039.7, 7094.2, 7289.0, 748.21, 1117.1, 1055.5, 504.95
40 Hz 7 days soluble	1:50	476.71, 283.83, 336.87, 237.22, 7.0175, 4.1480, 4.0580, 1.5205, 91.864, 152.73, 148.84, 141.07, 162.44	419.7, 248.1, 242.7, 90.974, 95.626, 56.936, 67.577, 47.586, 200.87, 13.56, 9.794, 15.44, 3.677
40 Hz 7 days insoluble	1:100	281.97, 270.37, 86.199, 239.71, 23.557, 15.166, 22.714, 1038.9, 1099.8, 1760.8, 1558.8, 187.69, 22.64	202.96, 130.71, 195.73, 193.70, 1646.89, 1579.1, 503.44, 1400.0, 7536.62, 955.23, 1208.8, 694.57, 784.91
Dark 1 hr BC	1:2	81.874, 18.343, 86.554	391.95, 883.69, 604.97
40 Hz 1 hr BC	1:2	81.307, 27.986, 30.113	300.34, 1152.5, 616.92
40 Hz 1 hr wait 4 hr	1:2	91.06, 141.8, 111.2, 12.30	108.0, 168.1, 157.3, 35.158
40 Hz 1 hr wait 12 hr	1:2	167.2, 101.6, 89.31, 119.9	236.1, 134.6, 124.8, 152.4
40 Hz 1 hr wait 24 hr	1:2	246.7, 177.6, 281.2, 175.0, 257.3, 204.2	231.8, 107.0, 402.7, 184.6, 245.1, 179.7
Dark APP/PS1	1:2	1050.16, 1085.25, 1522.45, 1153.69, 1750.77	19.22, 30.68, 28.08, 14.25, 25.30
40 Hz APP/PS1	1:2	512.42, 947.80, 850.45, 793.63	18.85, 15.58, 18.92, 11.44
Dark WT	1:1	0.038, 0.813, 2.016, 1.913, 0.313, 4.11, 7.23, 20.2, 40.4, 38.7, 11.9	N/A
40 Hz WT	1:1	0.139, 0.325, 0.346, 0.390, 8.92, 12.1, 6.34, 12.4, 13.1	N/A

Raw A β_{1-40} and A β_{1-42} levels with ELISA dilution ratios for each experimental group. Equal tissue masses were compared for each ELISA experiment. For 7-day experiments, values were normalized to within litter controls such that raw values 1–4 in each condition were normalized to the mean of 'Dark' values 1–4; raw values 5–9 in each condition were normalized to the mean of 'Dark' values 5–9; raw values 10–13 in each condition were normalized to the mean of 'Dark' values 10–13.

S-2-hydroxyglutarate regulates CD8⁺ T-lymphocyte fate

Petros A. Tyrakis^{1,2}, Asis Palazon¹, David Macias¹, Kian. L. Lee³, Anthony. T. Phan⁴, Pedro Veliça⁵, Jia You³, Grace S. Chia³, Jingwei Sim¹, Andrew Doedens⁴, Alice Abelanet¹, Colin E. Evans¹, John R. Griffiths², Lorenz Poellinger^{3,5}, Ananda W. Goldrath⁴ & Randall S. Johnson^{1,5}

R-2-hydroxyglutarate accumulates to millimolar levels in cancer cells with gain-of-function isocitrate dehydrogenase 1/2 mutations. These levels of R-2-hydroxyglutarate affect 2-oxoglutarate-dependent dioxygenases. Both metabolite enantiomers, R- and S-2-hydroxyglutarate, are detectable in healthy individuals, yet their physiological function remains elusive. Here we show that 2-hydroxyglutarate accumulates in mouse CD8⁺ T cells in response to T-cell receptor triggering, and accumulates to millimolar levels in physiological oxygen conditions through a hypoxia-inducible factor 1- α (HIF-1 α)-dependent mechanism. S-2-hydroxyglutarate predominates over R-2-hydroxyglutarate in activated T cells, and we demonstrate alterations in markers of CD8⁺ T-cell differentiation in response to this metabolite. Modulation of histone and DNA demethylation, as well as HIF-1 α stability, mediate these effects. S-2-hydroxyglutarate treatment greatly enhances the *in vivo* proliferation, persistence and anti-tumour capacity of adoptively transferred CD8⁺ T cells. Thus, S-2-hydroxyglutarate acts as an immunometabolite that links environmental context, through a metabolic-epigenetic axis, to immune fate and function.

In response to T-cell receptor (TCR) triggering, quiescent CD8⁺ T lymphocytes transition to a proliferative effector state. During this response, memory CD8⁺ T lymphocytes form and can persist for the entire lifespan of the organism, mounting rapid recall responses, thereby providing long-term immunity. The metabolic programs of these different CD8⁺ T-lymphocyte states are distinct and important for function^{1–4}. Effector CD8⁺ T lymphocytes generate most ATP and biomass through glycolysis⁵; both naive and memory cells rely heavily on oxidative phosphorylation^{6,7}. Various cytokines and transcription factors are important for the differentiation of CD8⁺ T lymphocytes, and it is evident that immunological memory is influenced by epigenetic mechanisms^{8–12}.

CD8⁺ T lymphocytes traffic into severely hypoxic areas within tumours and inflammatory tissue¹³. The response to oxygenation, mediated by the von Hippel-Lindau (VHL) and HIF- α proteins, is an essential regulator of metabolism and CD8⁺ T-lymphocyte function^{14–16}. Here we demonstrate that CD8⁺ T lymphocytes produce 2-hydroxyglutarate (2HG) in response to TCR triggering and environmental hypoxia. Using CD8⁺ T-lymphocytes isolated from *Mus musculus* with CD8-specific genetic deletions of Vhl, Hif-1 α and Hif-2 α , driven by *cre* expressed under the distal promoter of the lymphocyte protein tyrosine kinase (*dLck^{cre}*), we highlight the dependency of this metabolic feature on the HIF pathway. S-2HG constitutes the majority of the 2HG pool, and we show that S-2HG alters the phenotypic and functional characteristics of CD8⁺ T lymphocytes, maintaining a state of increased proliferative, survival and anti-tumour capacity.

The VHL–HIF-1 α axis regulates 2HG production

To elucidate the metabolic effects of HIF-1 α activation, we profiled the metabolome of CD8⁺ T lymphocytes with low (*Vhl^{fl/fl}*), or high (*Vhl^{fl/fl}dLck^{cre}*, denoted as *Vhl^{-/-}*) HIF signalling, and knockout of both Hif-1 α and Vhl (*Hif1a^{fl/fl}Vhl^{fl/fl}dLck^{cre}*, denoted as *Hif1a^{-/-}Vhl^{-/-}*) to control for a specific contribution of HIF-1 α ¹⁴. Unsupervised clustering

and principal component analysis (Fig. 1a, b) separate *Vhl^{-/-}* from *Vhl^{fl/fl}* CD8⁺ T lymphocytes. *Vhl^{-/-}Hif1a^{-/-}* cluster with *Vhl^{fl/fl}*, indicating that Hif-1 α mediates significant metabolic changes following *Vhl* deletion. Glycolysis is important for sustaining effector function^{1–3,17} and these data indicate that Vhl suppresses glycolysis via inhibition of Hif-1 α , (Extended Data Fig. 1a–c).

Vhl loss suppresses late and increases early tricarboxylic acid (TCA) cycle intermediates (Extended Data Fig. 1a). Notably, 2HG is significantly enriched in *Vhl^{-/-}* CD8⁺ T lymphocytes (Fig. 1c, Extended Data Fig. 1d). Furthermore, increases in 2HG are dependent on Hif-1 α when Vhl is deleted (Extended Data Fig. 1a). This was validated using quantitative mass spectrometry in *Vhl^{-/-}* and *Vhl^{-/-}Hif1a^{-/-}* CD8⁺ T lymphocytes (Fig. 1d), as well as in VHL-null cell lines, that express either HIF-1 α (RCC4) or HIF-2 α (786-O), reconstituted with VHL (Fig. 1e, Extended Data Fig. 1e). Deletion of *Vhl* in murine embryonic fibroblasts from *Vhl^{fl/fl}* mice increases 2HG levels (Fig. 1f, Extended Data Fig. 1f). Hence, the VHL–HIF signalling axis regulates 2HG levels, and constitutive Hif-1 α signalling underlies this effect in *Vhl*-null CD8⁺ T lymphocytes.

HIF-1 α regulates S-2HG production

R-2HG production is increased by isocitrate dehydrogenase 1 and/or isocitrate dehydrogenase 2 mutations in different cancers^{18–20}; accumulation of the S-2HG enantiomer occurs in the context of hypoxia^{21,22} and mitochondrial dysfunction^{23,24}. We thus sought to determine 2HG levels in CD8⁺ T lymphocytes following activation. 2HG before activation and at sea-level oxygen is undetectable, whereas levels at the same oxygenation are elevated 2–4 days after TCR stimulation (Fig. 2a). When activated CD8⁺ T lymphocytes are exposed to 1% oxygen, the intracellular concentration of 2HG reaches millimolar levels (Fig. 2b, Extended Data Fig. 2a) and is proportional to the degree of oxygenation (Fig. 2c). We confirmed this using ¹H NMR spectroscopy (Extended Data Fig. 2b). Given such high levels of 2HG, we sequenced²⁵

¹Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK. ²Cancer Research UK, Cambridge Institute, University of Cambridge, Cambridge, UK.

³Cancer Science Institute of Singapore, National University of Singapore, Singapore. ⁴Molecular Biology Section, UC San Diego, La Jolla, California, USA. ⁵Department of Cell and Molecular Biology, Karolinska Institute, Stockholm, Sweden.

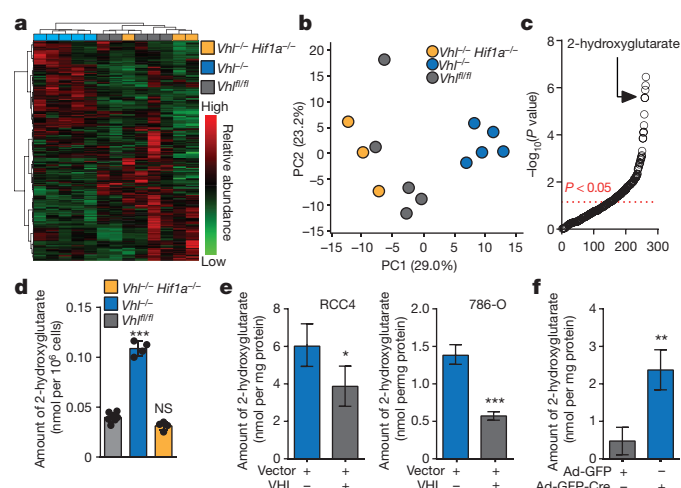


Figure 1 | VHL-HIF signalling regulates 2-hydroxyglutarate levels.

a, Unsupervised hierarchical clustering and heat map of all detected metabolites. **b**, PCA of metabolomes. Percentage variance of each PC is in parentheses. **c**, Metabolites ranked in order of decreasing *P* value. **d**, 2HG levels in *Vhl^{fl/fl}* (*n* = 7), *Vhl^{fl/fl} dLck^{cre}* (*n* = 4) and *Hif1a^{fl/fl} Vhl^{fl/fl} dLck^{cre}* (*n* = 4) CD8⁺ T lymphocytes activated with anti-CD3 and anti-CD28 antibodies and then cultured in IL-2 for a further 5 days. Each dot represents a mouse in **b** and **d**. **e**, 2HG levels in RCC4 and 786-O cells with reconstitution of VHL; *n* = 3. **f**, 2HG levels in MEFs with deletion of *Vhl*; *n* = 3. Unpaired *t*-test (**e**, **f**) and one-way ANOVA (**d**). Error bars denote s.d.; NS, not significant; **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

Ildh1/2, to preclude the unlikely possibility that culturing primary murine CD8⁺ T lymphocytes in hypoxia gives rise to mutations known to cause R-2HG production in humans¹⁸; we found no evidence for such mutations (Extended Data Fig. 3a–e). Resolving the enantiomers of 2HG indicates that S-2HG constitutes the majority of the 2HG pool (Fig. 2d). Primary human CD8⁺ T lymphocytes also accumulate S-2HG in hypoxia (Fig. 2e, f), indicating that this is not limited to mouse lymphocytes.

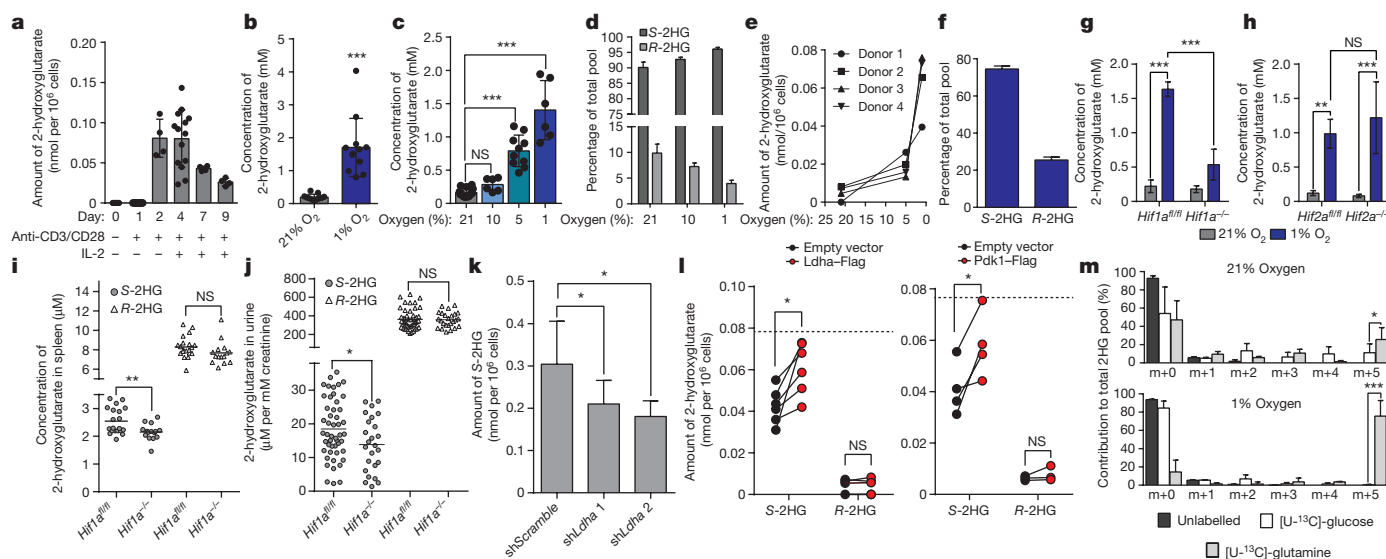


Figure 2 | Hypoxic induction of 2-hydroxyglutarate depends on Hif-1α in CD8⁺ T lymphocytes. **a**, 2HG in naive and activated cells; *n* ≥ 4 mice per time point. **b**, 2HG in cells cultured at 21% or 1% oxygen for 48 h; *n* = 12 mice. **c**, 2HG in cells cultured at 21% (*n* = 15), 10% (*n* = 6), 5% (*n* = 10) or 1% (*n* = 6) oxygen for 48 h. **d**, Enantioselective mass spectrometry (MS) for S- and R-2HG. **e**, 2HG in human cells (*n* = 4 healthy donors). **f**, Enantioselective MS for S- and R-2HG from cells in **e**. **g**, **h**, 2HG in *Hif1a^{fl/fl}* and *Hif1a^{fl/fl} dLck^{cre}* (**g**) or *Hif2a^{fl/fl}* and *Hif2a^{fl/fl} dLck^{cre}* (**h**) cells, cultured as in **b** (*n* = 4 mice per genotype). **i**, 2HG in splenic extracts from *Hif1a^{fl/fl}*

We carried out deletion of *loxP*-flanked *Hif1a* or *Hif2a* (also known as *Epas1*) genes in CD8⁺ T lymphocytes, using *dLck^{cre}* (ref. 26; Extended Data Fig. 2c). 2HG accumulation is abolished in *Hif1a^{fl/fl} dLck^{cre}* (denoted as *Hif1a^{-/-}*), but not *Hif2a^{fl/fl} dLck^{cre}* (denoted as *Hif2a^{-/-}*) cells under hypoxia (Fig. 2g, h; Extended Data Fig. 2d, e), with no difference in viability (Extended Data Fig. 2f). We next examined 2HG levels *in vivo*, in the spleens of mice. There is more R-2HG than S-2HG (Fig. 2i); furthermore, the levels of S-2HG are significantly decreased in spleens of *Hif1a^{fl/fl} dLck^{cre}* mice (Fig. 2i). There is also a slight decrease in the levels of R-2HG (Fig. 2i). 2HG is present in the urine of healthy individuals and is elevated in patients with 2HG acidurias²⁷. *Hif1a^{fl/fl} dLck^{cre}* mice have lower levels of S-2HG in urine (Fig. 2j) indicating that Hif-1α in the T-lymphocyte (CD4⁺ and CD8⁺) compartment makes a contribution to S-2HG production *in vivo*. Activated *Hif1a^{-/-}* CD8⁺ T lymphocytes in 21% oxygen have lower 2HG at extended time points (Extended Data Fig. 2g), indicating a contribution of Hif-1α in non-hypoxic conditions also.

We next sought to determine the metabolic route by which Hif-1α promotes S-2HG production in CD8⁺ T lymphocytes. Transcriptionally, hypoxic CD8⁺ T lymphocytes show induction of glycolysis and suppression of the TCA cycle (Extended Data Fig. 2h). Moreover, TCA cycle intermediates are decreased (Extended Data Fig. 2i). Recent reports implicate lactate and malate dehydrogenases (*Ldha* and *Mdh1/2*) as enzymatic sources of S-2HG in hypoxia^{21,22}. In *Hif1a^{-/-}* CD8⁺ T lymphocytes, the hypoxic expression of these enzymes suggests that *Mdh1* and *Mdh2* are unlikely to mediate the hypoxia-induced accumulation of S-2HG (Extended Data Fig. 2j). Confirming this, knockdown of *Mdh1* or *Mdh2* does not decrease S-2HG in hypoxia (Extended Data Fig. 2k–m); knockdown of *Mdh1* leads to marginal increases in S-2HG (Extended Data Fig. 2l). Knockdown of *Mdh1* or *Mdh2* increases R-2HG levels (Extended Data Fig. 2l, m). Knockdown of *Ldha* (Extended Data Fig. 2k) decreases S-2HG (Fig. 2k), and also increases R-2HG levels in hypoxic CD8⁺ T lymphocytes (Extended Data Fig. 2n). Overexpression of *Ldha* in hypoxic *Hif1a^{-/-}* CD8⁺ T lymphocytes (Extended Data Fig. 2o) rescues S-2HG production (Fig. 2l). Consistent with this, *Ldha*

(*n* = 20) and *Hif1a^{fl/fl} dLck^{cre}* (*n* = 15) mice. **j**, 2HG in urine from *Hif1a^{fl/fl}* (*n* = 49) and *Hif1a^{fl/fl} dLck^{cre}* (*n* = 25) mice. **k**, S-2HG in cells with sh*Ldha* (*n* = 4 individual transductions). **l**, *Ldha*-Flag (*n* = 6) or *Pdk1*-Flag (*n* = 4) overexpression in *Hif1a^{fl/fl} dLck^{cre}* cells. The dotted line represents S-2HG levels in *Hif1a^{fl/fl}* cells (*n* = 4/6 individual transductions). **m**, ¹³C-isotopologue profile of 2HG in cells cultured as in **b** (*n* = 7 mice per condition). Unpaired *t*-test (**i**, **j**). Paired *t*-test (**b**, **l**, **m**). One-way ANOVA (**c**, **k**) and two-way ANOVA (**g**, **h**). Error bars denote s.d.; each dot in **a**–**c**, **i** and **j** represents a mouse. **P* < 0.05, ***P* < 0.01, ****P* < 0.001; NS, not significant.

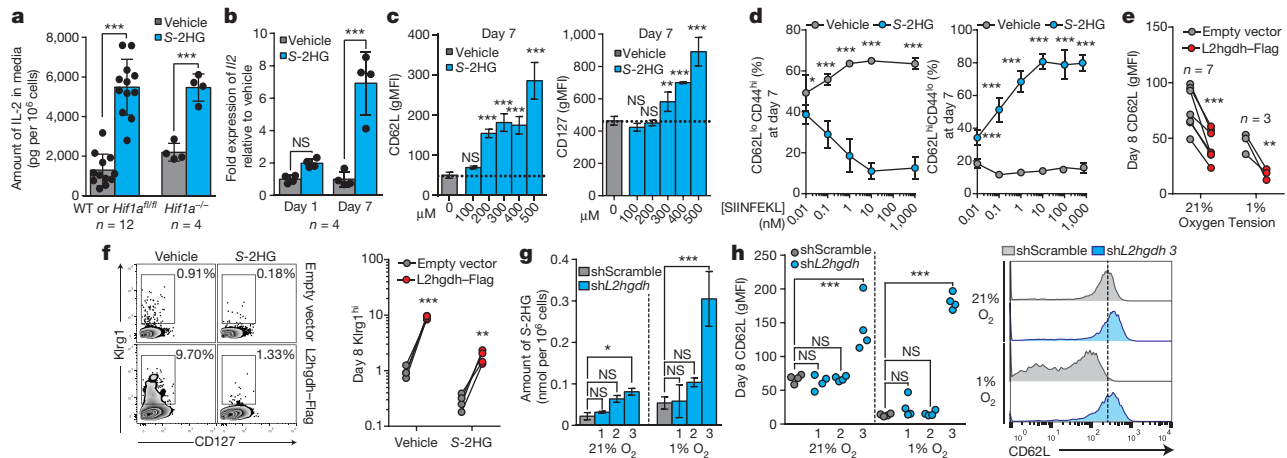


Figure 3 | S-2HG alters phenotypic marker expression of CD8⁺ T lymphocytes. **a**, Media IL-2 levels from S-2HG-treated cells. **b**, Expression of *Il2* mRNA in S-2HG-treated cells. **c**, CD62L and CD127 expression on S-2HG-treated OT-I cells ($n = 4$ mice). **d**, Percentage of CD62L^{hi}CD44^{lo} and percentage of CD62L^{lo}CD44^{hi} cells with S-2HG treatment ($n = 4$ mice). **e**, CD62L expression on cells treated with S-2HG and transduced with empty or L2hgdh-Flag over-expression vectors. **f**, Representative flow cytometry plots of Klr1 compared with CD127 ($n = 4$) on cells from **e**.

Associated statistics are shown. **g**, S-2HG level in response to shRNA against *L2hgdh* (shL2hgdh) ($n = 4$). **h**, CD62L expression in response to *L2hgdh* knockdown ($n = 4$). Representative flow cytometry histogram of CD62L surface levels is shown on the right. Two-way ANOVA (**a**, **b**, **e**, **f**). One-way ANOVA (**c**, **d**, **g**, **h**). Error bars denote s.d. Each dot in **a**, **b**, **e**, **f**, **h**, represents a mouse. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; NS, not significant. gMFI, geometric mean fluorescence intensity.

expression in hypoxic CD8⁺ T lymphocytes is HIF-1 α -dependent (Extended Data Fig. 2j).

We next performed ¹³C-labelling experiments and, in agreement with previous reports^{21,23}, uniformly labelled (U-¹³C)-glucose and U-¹³C-glutamine tracing indicates that glutamine is the major source of 2HG, (Fig. 2m, Extended Data Fig. 4a–c). The m+5 isotopologue dominates, suggesting direct conversion of glutamine-derived 2-oxoglutarate to 2HG^{21,23}. Furthermore, the glutamate pool increases in hypoxic conditions (Extended Data Fig. 2p) and in *Vhl*^{-/-} CD8⁺ T lymphocytes (Extended Data Fig. 1a) and depends on Hif-1 α (Extended Data Fig. 2q) but not Hif-2 α (Extended Data Fig. 2r). Inhibition of pyruvate dehydrogenase (Pdh) by pyruvate dehydrogenase kinases (Pdk), promotes glutaminolysis^{28–30}. We reasoned that Pdk supports S-2HG production by diverting glutamine-derived 2-oxoglutarate to Ldha and, consistent with this, dichloroacetate (DCA) abrogates hypoxia-induced 2HG accumulation (Extended Data Fig. 2s–u). Pdk1 expression is impaired in hypoxic *Hif1a*^{-/-} CD8⁺ T lymphocytes (Extended Data Fig. 2j), as is phosphorylation of Pdh (Extended Data Fig. 2s), and re-expression of Pdk1 in this context increases S-2HG in hypoxia (Fig. 2l, Extended Data Fig. 2o). Inhibition of Pdk activity also impedes hypoxia-induced increases in the glutamate pool (Extended Data Fig. 2v). Hence, HIF-1 α drives S-2HG production in hypoxic CD8⁺ T lymphocytes through the Pdk–Pdh signalling axis and Ldha induction (Extended Data Fig. 4a).

S-2HG alters CD8⁺ T-cell differentiation

S-2HG inhibits 2-oxoglutarate-dependent dioxygenases^{31,32}. Consistent with this, Hif-1 α is stabilized in normoxic (Extended Data Fig. 5a) and hypoxic (Extended Data Fig. 5b) CD8⁺ T lymphocytes by treatment with cell-permeable S-2HG, suggesting that S-2HG augments HIF signalling in normoxia and hypoxia. Additionally, there is increased phosphorylation of Pdh-E1 α (Extended Data Fig. 5a, b), elevated glucose uptake, lactate secretion (Extended Data Fig. 5c) and Vegfa production (Extended Data Fig. 5d), indicating Hif-1 α -dependent effects. As Hif-1 α supports effector functions in CD8⁺ T lymphocytes^{13,15}, we reasoned that S-2HG promotes effector differentiation through Hif-1 α . However, unexpectedly, there is suppression of effector cytokine production (Extended Data Fig. 5e) and decreased cytotoxicity (Extended Data Fig. 5f). Furthermore, S-2HG restrains cell expansion (Extended Data Fig. 5g) and there is a clear

increase in apoptosis at doses greater than 300 μ M (Extended Data Fig. 5h, i). Further characterization revealed decreased secretion of IFN- γ (Extended Data Fig. 5j), yet elevated production of IL-2 (Fig. 3a), with increased viability in the absence of IL-2 supplementation (Extended Data Fig. 5k). This possibly reflects an autocrine pro-survival effect. These effects are robustly mediated at the transcriptional level after prolonged treatment with S-2HG (Fig. 3b, Extended Data Fig. 5l) and are independent of HIF-1 α (Fig. 3a, Extended Data Fig. 5j, k).

We then characterized the phenotype of cells that had been treated with a prolonged course of S-2HG. There is increased expression of CD62L (Fig. 3c, Extended Data Fig. 5m) and this is reversible upon withdrawal of treatment (Extended Data Fig. 5n). The effect does not occur when treating cells cultured in vehicle for 7 days (Extended Data Fig. 5n), demonstrating that S-2HG treatment of newly activated cells maintains this phenotypic marker. Importantly, CD62L downregulation does not occur when Hif-1 α is absent¹⁵, which masks the effect of S-2HG on CD62L following Hif-1 α deletion (Extended Data Fig. 5o). Hif-2 α is dispensable for CD62L maintenance in response to S-2HG (Extended Data Fig. 5p). With S-2HG, CD62L maintenance depends on the level of antigenic stimulation (Fig. 3d). Furthermore, S-2HG-treated cells express more CD127 (Fig. 3c), CD44, 41BB, Eomes and less PD-1 in a Hif-1 α -independent manner (Extended Data Fig. 5q).

To determine the role of endogenously produced S-2HG, overexpression of L2hgdh (Extended Data Fig. 5r), a dehydrogenase that oxidizes S-2HG, was performed. Overexpression of L2hgdh promotes the downregulation of CD62L in both 21% and 1% oxygen (Fig. 3e), indicating that endogenously produced S-2HG regulates CD62L expression. Furthermore, L2hgdh overexpression leads to an increase in the proportion of Klr1^{hi} cells, which are decreased in the presence of exogenous S-2HG (Fig. 3f). Conversely, successful short hairpin RNA (shRNA)-mediated knockdown of *L2hgdh* by hairpin 3 (Extended Data Fig. 5s) increases endogenous S-2HG levels (Fig. 3g), especially in 1% oxygen, promoting maintenance of CD62L (Fig. 3h). Knockdown of *L2hgdh* blocks loss of CD62L in response to low oxygen (Fig. 3h). The same effect is seen with CD127 in low oxygen (Extended Data Fig. 5t). These data demonstrate that L2hgdh activity regulates the expression of key phenotypic markers of CD8⁺ T lymphocytes, by controlling endogenous S-2HG levels. Transcriptionally, S-2HG treatment increases expression of *Eomes*, *Ccr6*, *Bcl6*, *Sell* (*Cd62L*) and *Tcf7*,

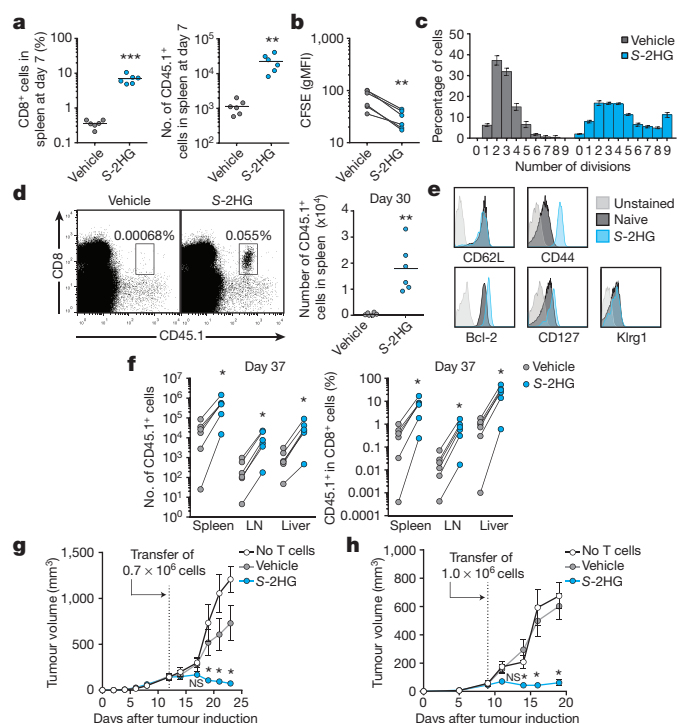


Figure 4 | S-2HG treatment promotes *in vivo* homeostatic renewal, persistence and anti-tumour capacity of transferred cells. **a**, Recovery of co-transferred CD45.1⁺ OT-I cells, from spleens ($n = 6$). **b**, *In vivo* CFSE levels in cells from **a**. **c**, Percentage of cells in **a** that divided 0–9 times *in vivo*. **d**, Representative flow cytometry plots and associated statistics of recovered cells from spleens ($n = 6$). **e**, Representative analysis of cells in **d** relative to naive cells ($n = 6$). **f**, Recovery of co-transferred cells, from spleens, lymph nodes and livers of vaccinated mice ($n = 6$). **g**, Lymphocyte-depleted mice bearing EG7-OVA tumours treated with no T cells ($n = 7$) or OT-I cells cultured with or without ($n = 6$) S-2HG. Error bars denote s.e.m. **h**, Lymphoreplete mice bearing EG7-OVA tumours treated with no T cells or OT-I cells cultured with or without S-2HG ($n = 6$). Error bars denote s.e.m. Paired *t*-test (**a**, **b**), unpaired *t*-test (**d**), one-way ANOVA (**g**, **h**). Error bars denote s.d. (not in **g** and **h**). Each dot in **a**, **b**, **d** and **f** represents a mouse. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; NS, not significant.

with repression of *Prdm1*, after 7 days (Extended Data Fig. 5u). This transcriptional program is similar to gene expression changes in memory CD8⁺ T lymphocytes, suggesting that S-2HG treatment of CD8⁺ T lymphocytes *ex vivo* may enhance long-term persistence and survival in the context of adoptive cell transfer³³.

We thus co-transferred CFSE-labelled vehicle and S-2HG treated CD45.1.1 or CD45.1.2 OT-I CD8⁺ T lymphocytes into lymphocyte-depleted mice (Extended Data Fig. 6a) to assess their capacity for homeostatic proliferation^{34,35}. S-2HG-treated cells display greater homeostatic proliferation (Fig. 4a, b), with more cells dividing more than 5 times (Fig. 4c). We then assessed the capacity of S-2HG treated cells to persist in lymphoreplete mice. Adoptively transferred ovalbumin-specific CD45.1 OT-I CD8⁺ T lymphocytes, pre-treated with S-2HG, show markedly enhanced persistence 30 days after transfer (Fig. 4d), expressing elevated CD44, CD127 and Bcl-2 levels relative to naive cells^{36,37} (Fig. 4e). In response to vaccination with ovalbumin-derived protein SIINFEKL-loaded dendritic cells, S-2HG-treated OT-I CD8⁺ T lymphocytes robustly recall (Fig. 4f and Extended Data Fig. 6b, c). Consistent with this, OT-I CD8⁺ T lymphocytes, pre-treated with S-2HG are more proficient at controlling tumour growth *in vivo* in both lymphocyte-depleted (Fig. 4g) and lymphoreplete (Fig. 4h) mice. These data demonstrate that S-2HG treatment *ex vivo* maintains cells in a state with increased proliferative and survival capacity, when transferred *in vivo*, that is otherwise decreased by effector differentiation.

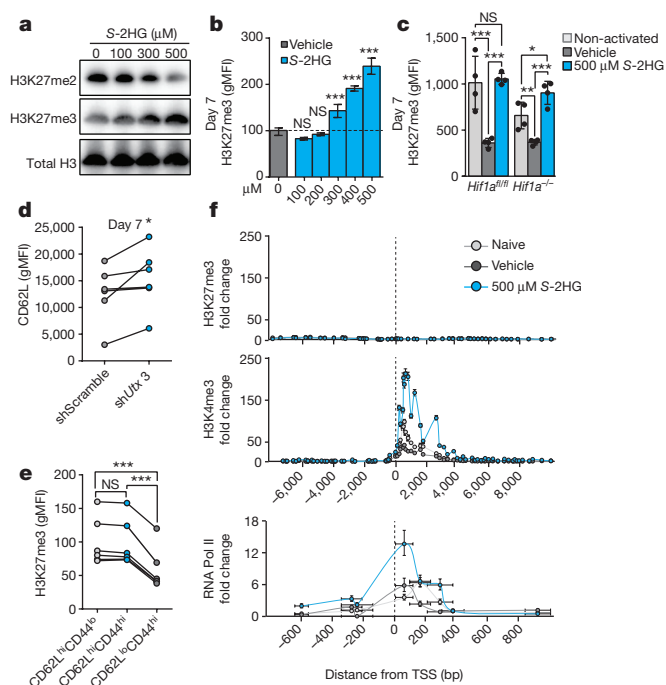


Figure 5 | S-2HG alters global H3K27me3 in CD8⁺ T lymphocytes. **a**, Immunoblot analysis for H3K27me2 and H3K27me3 in cells treated with S-2HG. **b**, H3K27me3 in OT-I cells treated with S-2HG ($n = 6$ mice). **c**, H3K27me3 abundance in *Hif1a*^{fl/fl} and *Hif1a*^{fl/fl} *dLck*^{cre} cells treated with S-2HG ($n = 4$). **d**, CD62L expression by cells with *shUtx 3*. **e**, *In vivo* H3K27me3 levels in CD8⁺ populations ($n = 6$ mice). **f**, ChIP-qPCR for H3K4me3, H3K27me and RNA Pol II around the TSS for CD62L. A pool of $n = 6$ mice was used; error bars denote s.e.m. One-way ANOVA (**b**, **c**). Error bars denote s.d. (**b**, **c**), and each dot in **c** and **e** represents a mouse. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; NS, not significant. For immunoblot source images, see Supplementary Fig. 1. gMFI, geometric mean fluorescence intensity.

S-2HG alters methylation in CD8⁺ T cells

Mechanistic target of rapamycin (mTOR) is a modifier of CD8⁺ T-lymphocyte differentiation; however, we do not observe mTOR inhibition^{36,38} at doses of S-2HG needed for this to occur (Extended Data Fig. 7). S-2HG treatment *ex vivo* may be selecting cells that express higher levels of anti-apoptotic genes. Two critical anti-apoptotic genes implicated in CD8⁺ T-lymphocyte survival are Bcl-2 and Bcl-XL^{39,40}. These genes are not induced by S-2HG treatment (Extended Data Fig. 8a–d). Moreover, overexpression of Bcl-2 or Bcl-XL (Extended Data Fig. 8e) does not influence the expression of CD62L, CD44 or CD127 in the presence or absence of S-2HG (Extended Data Fig. 8f–i), indicating that S-2HG is exerting these phenotypic changes independently of Bcl-2 or Bcl-XL.

Inhibition of 2-oxoglutarate-dependent dioxygenases that demethylate histones (Jumonji C containing proteins) or oxidise 5-methylcytosine in DNA (Ten-eleven translocation (Tet) proteins) may mediate the effect of S-2HG^{32,41–43}. S-2HG alters global levels of various histone methylation marks (Extended Data Fig. 9a); in particular, di- and tri-methylation on H3K27 are reciprocally altered, indicating inhibition of H3K27me3 demethylation (Fig. 5a, b, Extended Data Fig. 9b). The H3K27me2/3 demethylase Utx (also known as Kdm6a) is an important regulator of thymocyte differentiation⁴⁴; changes in H3K27me3 levels correlate with genes associated with regulation of CD8⁺ T lymphocyte differentiation¹¹. Global levels of H3K27me3 in CD8⁺ T lymphocytes are reduced following activation, but remain high with S-2HG treatment in a HIF-1 α -independent manner (Fig. 5c, Extended Data Fig. 9c). Utx is induced following TCR stimulation (Extended Data Fig. 9d) and inhibition of Utx reproduces the effect of S-2HG treatment on CD62L expression (Fig. 5d, Extended Data Fig. 9e, f). *In vivo*, levels

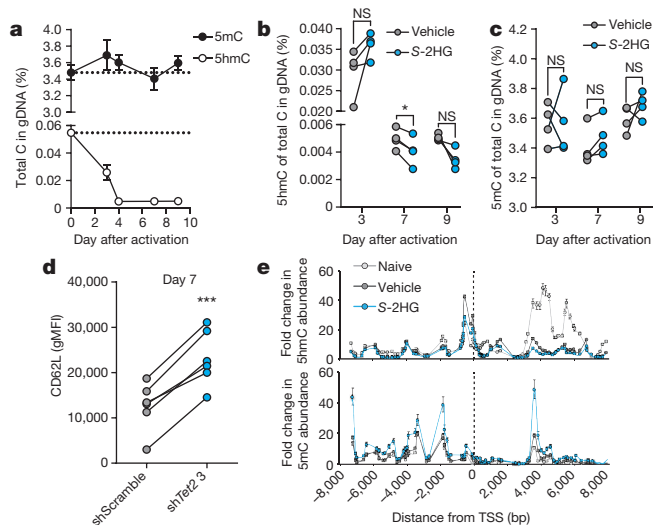


Figure 6 | S-2HG alters global 5hmC and 5mC in DNA of CD8⁺ T lymphocytes. **a**, 5hmC and 5mC in gDNA ($n = 4$ mice). **b**, 5hmC in gDNA from cells treated with S-2HG ($n = 4$ mice). **c**, 5mC in gDNA from cells treated with S-2HG ($n = 4$ mice). **d**, CD62L expression on CD8⁺ T lymphocytes 7 days after transduction with shTet2 3. **e**, DIP-qPCR for 5mC and 5hmC around the TSS for CD62L. A pool of $n = 6$ mice was used; error bars denote s.e.m. One-way ANOVA (**b** and **c**). Error bars denote s.d. (**a**) and each dot in **b** and **c** represents a mouse. * $P < 0.05$, *** $P < 0.001$; NS, not significant. gMFI, geometric mean fluorescence intensity.

of H3K27me3 are highest in central memory (CD62L^{hi}CD44^{hi}) and naive (CD62L^{hi}CD44^{lo}) CD8⁺ T lymphocytes, relative to effectors (CD62L^{lo}CD44^{hi}) (Fig. 5e). To determine whether histone methylation changes occur at the transcription start site (TSS) of CD62L with S-2HG treatment, we performed chromatin immunoprecipitation (ChIP)-PCR for H3K27me3, H3K4me3 and RNA polymerase II (RNA Pol II), on naive and activated CD8⁺ T lymphocytes cultured with or without S-2HG (Fig. 5f, Extended Data Fig. 9g). We could find no enrichment for H3K27me3 at or around the TSS. However, naive and S-2HG-treated CD8⁺ T lymphocytes have higher enrichment for H3K4me3 at the TSS that is reduced in vehicle-treated cells. Additionally, S-2HG treated CD8⁺ T lymphocytes have markedly higher RNA Pol II binding than both naive and vehicle-treated cells. Thus, S-2HG promotes CD62L transcription directly through enrichment of H3K4me3 at the TSS and indirectly through preservation of H3K27me3 elsewhere in the genome.

Total levels of 5-methylcytosine (5mC) in genomic DNA are largely unchanged by TCR triggering (Fig. 6a). However, total levels of 5-hydroxymethylcytosine (5hmC) decrease following TCR triggering (Fig. 6a). 5hmC removal in genomic DNA can occur through Tet-mediated oxidation (active) as well as DNA replication (passive)^{45,46}. S-2HG treatment induces small changes in total 5hmC and 5mC in a time-dependent manner (Fig. 6b, c); at day 3, there are marginally higher levels of 5hmC (statistically non-significant) with no changes in 5mC (Fig. 6b, c). This is probably due to decreased proliferation in the presence of S-2HG (Extended Data Fig. 5g). Following sustained treatment, at days 7 and 9, cells have less 5hmC and more 5mC (Fig. 6b, c). 5mC changes did not reach statistical significance at any time point; changes in 5hmC were statistically significant at day 7 only. Nonetheless, the changes at days 7 and 9 are consistent with inhibition of Tet proteins in the presence of sustained S-2HG treatment.

Tet2 regulates CD4⁺ T-lymphocyte function and is inhibited by 2HG^{47,48}. Knockdown of Tet2 recapitulates the effect of S-2HG on CD62L, indicating that Tet2 also contributes to CD8⁺ T-lymphocyte effector differentiation (Fig. 6d, Extended Data Fig. 10a), implicating DNA demethylation as an added epigenetic modifier of CD8⁺ T-cell

differentiation¹². We performed both 5hmC and 5mC DNA immunoprecipitation (DIP)-PCR around the TSS of CD62L (Fig. 6e, Extended Data Fig. 10b). Naive cells have the highest and lowest enrichment for 5hmC and 5mC respectively, whereas S-2HG-treated cells display the opposite pattern. Vehicle-treated cells have an intermediate level of these two marks.

Discussion

Our data support a model in which both Utx and Tet2 contribute to effector differentiation of CD8⁺ T lymphocytes *in vitro*. The activity of these epigenetic modifiers is altered by S-2HG in a fashion that inhibits effector differentiation.

TCR-triggering induced loss of 5hmC in genomic DNA, whereas 5mC levels were relatively stable (Fig. 6a). 5hmC levels stabilized by four days after activation (Fig. 6a) and S-2HG treatment produced reciprocal changes in 5mC and 5hmC levels at later time points (Fig. 6b, c). We investigated 5mC and 5hmC presence at and around the TSS of CD62L (Fig. 6e). S-2HG treatment causes reciprocal 5mC increases and 5hmC decreases in this region consistent with inhibition of Tet proteins and transcriptional repression⁴⁷, yet there is robust expression of CD62L with S-2HG treatment (Fig. 3c–e, h, Extended Data Fig. 5m–p). Inhibition of Tet2 maintains CD62L expression (Fig. 6d, Extended Data Fig. 10a), indicating that S-2HG-induced DNA methylation changes elsewhere in the genome may indirectly promote CD62L expression.

We also investigated H3K4me3 and H3K27me3 patterns at the TSS of CD62L (Fig. 5f). S-2HG-treated and naive CD8⁺ T lymphocytes have high enrichment for the active H3K4me3 mark⁴⁹ at the TSS that is lost in vehicle-treated cells. In S-2HG-treated cells, this is accompanied by binding of RNA Pol II. Despite global increases in H3K27me3 with S-2HG (Fig. 5a–c), the lack of H3K27me3 at the TSS of CD62L is not surprising, as this mark is associated with repression⁴⁹. Nevertheless, inhibition of Utx promotes CD62L maintenance (Fig. 5d, Extended Data Fig. 9e, f), indicating that H3K27me3 deposition at other genomic sites can indirectly promote CD62L expression. Owing to the relationship between histone methylation and other marks such as histone acetylation, modulation of the latter may have similar effects to those seen with S-2HG treatment⁵⁰.

Adoptively transferred cells treated with S-2HG *ex vivo* have an increased capacity to proliferate and persist *in vivo*, with enhanced anti-tumour efficacy (Fig. 4), demonstrating a new strategy to improve persistence of adoptive cell therapies for cancer. The data presented uncover a metabolic–epigenetic axis that controls aspects of T-cell fate. Factors regulating endogenous S-2HG levels, such as HIF signalling, TCR triggering, L2hdgh activity and potentially others, can alter the differentiation of CD8⁺ T lymphocytes and thus shape the immune response.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Data Availability The data that support the findings are available from the corresponding author upon reasonable request.

Received 14 August 2015; accepted 18 October 2016.

Published online 26 October 2016.

- Ho, P. C. *et al.* Phosphoenolpyruvate is a metabolic checkpoint of anti-tumor T cell responses. *Cell* **162**, 1217–1228 (2015).
- Chang, C. H. *et al.* Posttranscriptional control of T cell effector function by aerobic glycolysis. *Cell* **153**, 1239–1251 (2013).
- Chang, C. H. *et al.* Metabolic competition in the tumor microenvironment is a driver of cancer progression. *Cell* **162**, 1229–1241 (2015).
- Oestreich, K. J. *et al.* Bcl-6 directly represses the gene program of the glycolysis pathway. *Nat. Immunol.* **15**, 957–964 (2014).
- MacIver, N. J., Michalek, R. D. & Rathmell, J. C. Metabolic regulation of T lymphocytes. *Annu. Rev. Immunol.* **31**, 259–283 (2013).
- Pearce, E. L. *et al.* Enhancing CD8 T-cell memory by modulating fatty acid metabolism. *Nature* **460**, 103–107 (2009).

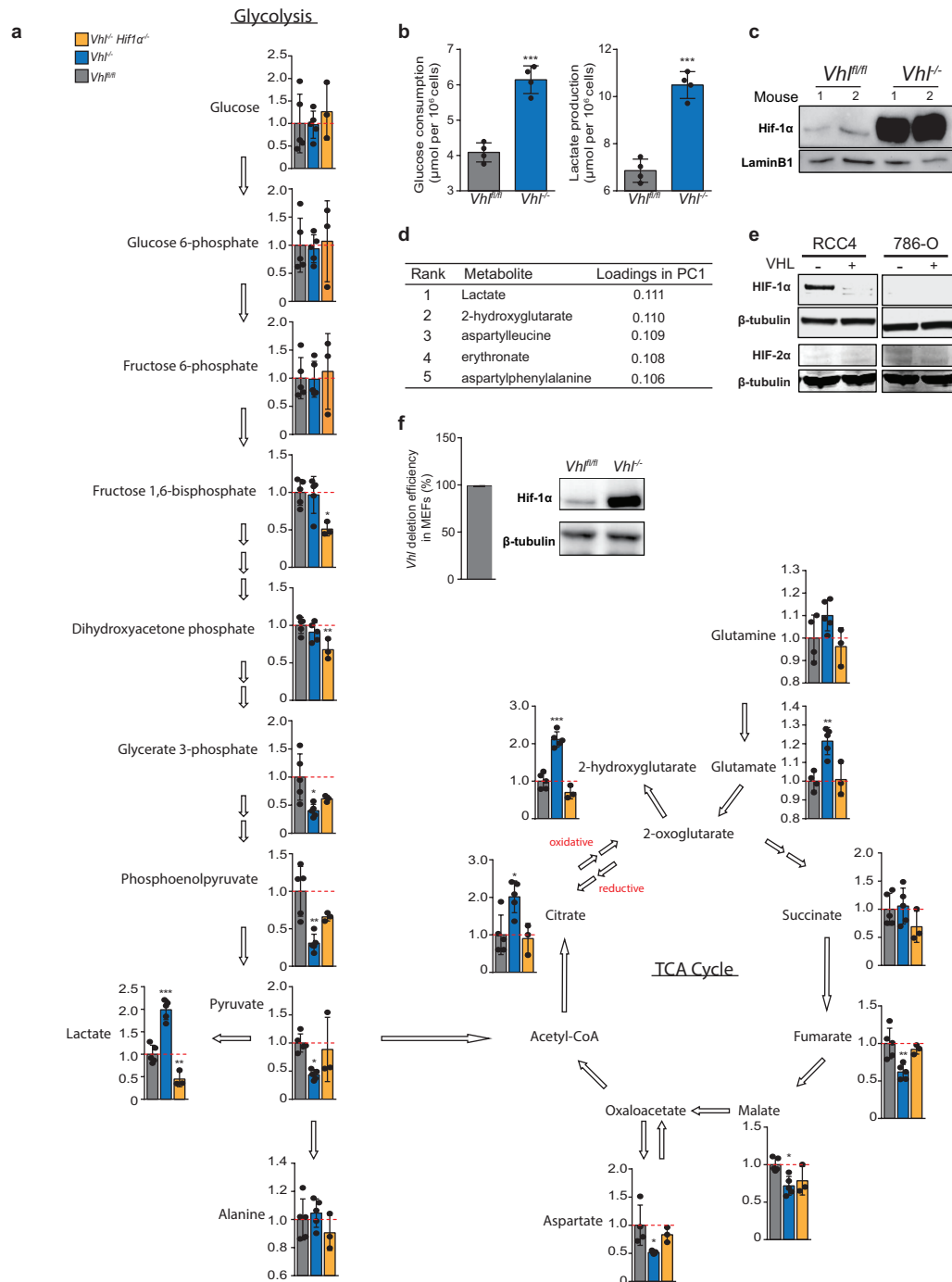
7. O'Sullivan, D. *et al.* Memory CD8⁺ T cells use cell-intrinsic lipolysis to support the metabolic programming necessary for development. *Immunity* **41**, 75–88 (2014).
8. Crompton, J. G. *et al.* Lineage relationship of CD8 T cell subsets is revealed by progressive changes in the epigenetic landscape. *Cell. Mol. Immunol.* (2015).
9. Harland, K. L. *et al.* Epigenetic plasticity of *Cd8a* locus during CD8⁺ T-cell development and effector differentiation and reprogramming. *Nat. Commun.* **5**, 3547 (2014).
10. Araki, Y. *et al.* Genome-wide analysis of histone methylation reveals chromatin state-based regulation of gene transcription and function of memory CD8⁺ T cells. *Immunity* **30**, 912–925 (2009).
11. Russ, B. E. *et al.* Distinct epigenetic signatures delineate transcriptional programs during virus-specific CD8⁺ T cell differentiation. *Immunity* **41**, 853–865 (2014).
12. Kaech, S. M. & Cui, W. Transcriptional control of effector and memory CD8⁺ T cell differentiation. *Nat. Rev. Immunol.* **12**, 749–761 (2012).
13. Palazon, A., Goldrath, A. W., Nizet, V. & Johnson, R. S. HIF transcription factors, inflammation, and immunity. *Immunity* **41**, 518–528 (2014).
14. Doedens, A. L. *et al.* Hypoxia-inducible factors enhance the effector responses of CD8⁺ T cells to persistent antigen. *Nat. Immunol.* **14**, 1173–1182 (2013).
15. Finlay, D. K. *et al.* PDK1 regulation of mTOR and hypoxia-inducible factor 1 integrate metabolism and migration of CD8⁺ T cells. *J. Exp. Med.* **209**, 2441–2453 (2012).
16. Clever, D. *et al.* Oxygen sensing by T cells establishes an immunologically tolerant metastatic niche. *Cell* **166**, 1117–1131 (2016).
17. Ghesquière, B., Wong, B. W., Kuchnio, A. & Carmeliet, P. Metabolism of stromal and immune cells in health and disease. *Nature* **511**, 167–176 (2014).
18. Losman, J. A. & Kaelin, W. G., Jr What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes Dev.* **27**, 836–852 (2013).
19. Saha, S. K. *et al.* Mutant IDH inhibits HNF-4 α to block hepatocyte differentiation and promote biliary cancer. *Nature* **513**, 110–114 (2014).
20. Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
21. Intlekofer, A. M. *et al.* Hypoxia induces production of L-2-hydroxyglutarate. *Cell Metab.* **22**, 304–311 (2015).
22. Oldham, W. M., Clish, C. B., Yang, Y. & Loscalzo, J. Hypoxia-mediated increases in L-2-hydroxyglutarate coordinate the metabolic response to reductive stress. *Cell Metab.* **22**, 291–303 (2015).
23. Wise, D. R. *et al.* Hypoxia promotes isocitrate dehydrogenase-dependent carboxylation of α -ketoglutarate to citrate to support cell growth and viability. *Proc. Natl Acad. Sci. USA* **108**, 19611–19616 (2011).
24. Mullen, A. R. *et al.* Oxidation of α -ketoglutarate is required for reductive carboxylation in cancer cells with mitochondrial defects. *Cell Reports* **7**, 1679–1690 (2014).
25. Patel, K. P. *et al.* Diagnostic testing for IDH1 and IDH2 variants in acute myeloid leukemia: an algorithmic approach using high-resolution melting curve analysis. *J. Mol. Diagn.* **13**, 678–686 (2011).
26. Zhang, D. J. *et al.* Selective expression of the Cre recombinase in late-stage thymocytes using the distal promoter of the *Lck* gene. *J. Immunol.* **174**, 6725–6731 (2005).
27. Gibson, K. M. *et al.* Stable-isotope dilution analysis of D- and L-2-hydroxyglutaric acid: application to the detection and prenatal diagnosis of D- and L-2-hydroxyglutaric acidemias. *Pediatr. Res.* **34**, 277–280 (1993).
28. Metallo, C. M. *et al.* Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. *Nature* **481**, 380–384 (2011).
29. Fendt, S. M. *et al.* Reductive glutamine metabolism is a function of the α -ketoglutarate to citrate ratio in cells. *Nat. Commun.* **4**, 2236 (2013).
30. Rajagopalan, K. N. *et al.* Metabolic plasticity maintains proliferation in pyruvate dehydrogenase deficient cells. *Cancer Metab.* **3**, 7 (2015).
31. Koivunen, P. *et al.* Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation. *Nature* **483**, 484–488 (2012).
32. Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30 (2011).
33. Arsenio, J. *et al.* Early specification of CD8⁺ T lymphocyte fates during adaptive immunity revealed by single-cell gene-expression analyses. *Nat. Immunol.* **15**, 365–372 (2014).
34. Goldrath, A. W. *et al.* Cytokine requirements for acute and basal homeostatic proliferation of naive and memory CD8⁺ T cells. *J. Exp. Med.* **195**, 1515–1522 (2002).
35. Murali-Krishna, K. *et al.* Persistence of memory CD8 T cells in MHC class I-deficient mice. *Science* **286**, 1377–1381 (1999).
36. Araki, K. *et al.* mTOR regulates memory CD8 T-cell differentiation. *Nature* **460**, 108–112 (2009).
37. Grayson, J. M., Zajac, A. J., Altman, J. D. & Ahmed, R. Cutting edge: increased expression of Bcl-2 in antigen-specific memory CD8⁺ T cells. *J. Immunol.* **164**, 3950–3954 (2000).
38. Fu, X. *et al.* 2-hydroxyglutarate inhibits ATP synthase and mTOR signaling. *Cell Metab.* **22**, 508–515 (2015).
39. Kurtulus, S. *et al.* Bcl-2 allows effector and memory CD8⁺ T cells to tolerate higher expression of Bim. *J. Immunol.* **186**, 5729–5737 (2011).
40. Song, A., Tang, X., Harms, K. M. & Croft, M. OX40 and Bcl-xl promote the persistence of CD8 T cells to recall tumor-associated antigen. *J. Immunol.* **175**, 3534–3541 (2005).
41. Ko, M. *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839–843 (2010).
42. Chowdhury, R. *et al.* The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* **12**, 463–469 (2011).
43. Lu, C. *et al.* IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474–478 (2012).
44. Manna, S. *et al.* Histone H3 Lysine 27 demethylases Jmjd3 and Utx are required for T-cell differentiation. *Nat. Commun.* **6**, 8152 (2015).
45. Rasmussen, K. D. & Helin, K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev.* **30**, 733–750 (2016).
46. Bachman, M. *et al.* 5-hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* **6**, 1049–1055 (2014).
47. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567 (2010).
48. Ichiyama, K. *et al.* The methylcytosine dioxygenase Tet2 promotes DNA demethylation and activation of cytokine gene expression in T cells. *Immunity* **42**, 613–626 (2015).
49. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
50. Zhang, F. *et al.* Epigenetic manipulation restores functions of defective CD8⁺ T cells from chronic viral infection. *Mol. Ther.* **22**, 1698–1706 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Cambridge Institute PK Bioanalytics Core facility for MS measurements; A. Cowburn, A. Wood, K. Lodge and E. Chilvers for human PBMCs; B. Jaggs for help with mouse work. P.A.T. was funded by CRUK, the MRC (1495954) and Wellcome Trust. J.R.G. was funded by CRUK. A.P. was funded by Marie-Curie IEF. A.T.P. was funded by UCSD NIH Grant (5T32GM007240-36). A.W.G. was funded by the NIH (A1096852, A1072117), Leukemia and Lymphoma Society, and Pew Scholars Fund. K.L.L., J.Y., G.S.C. and L.P. were funded by the Singapore National Research Foundation and Singapore Ministry of Education, the NMRC Clinician Scientist (NMRC/CIRG/1389/2014.) and the Swedish Research Council. R.S.J. and co-workers are funded by the Wellcome Trust (grant WT092738MA), the Swedish Cancer Foundation (Cancerfonden), and the Swedish Research Council (Vetenskapsrådet). This work is dedicated to the memory of Lorenz Poellinger.

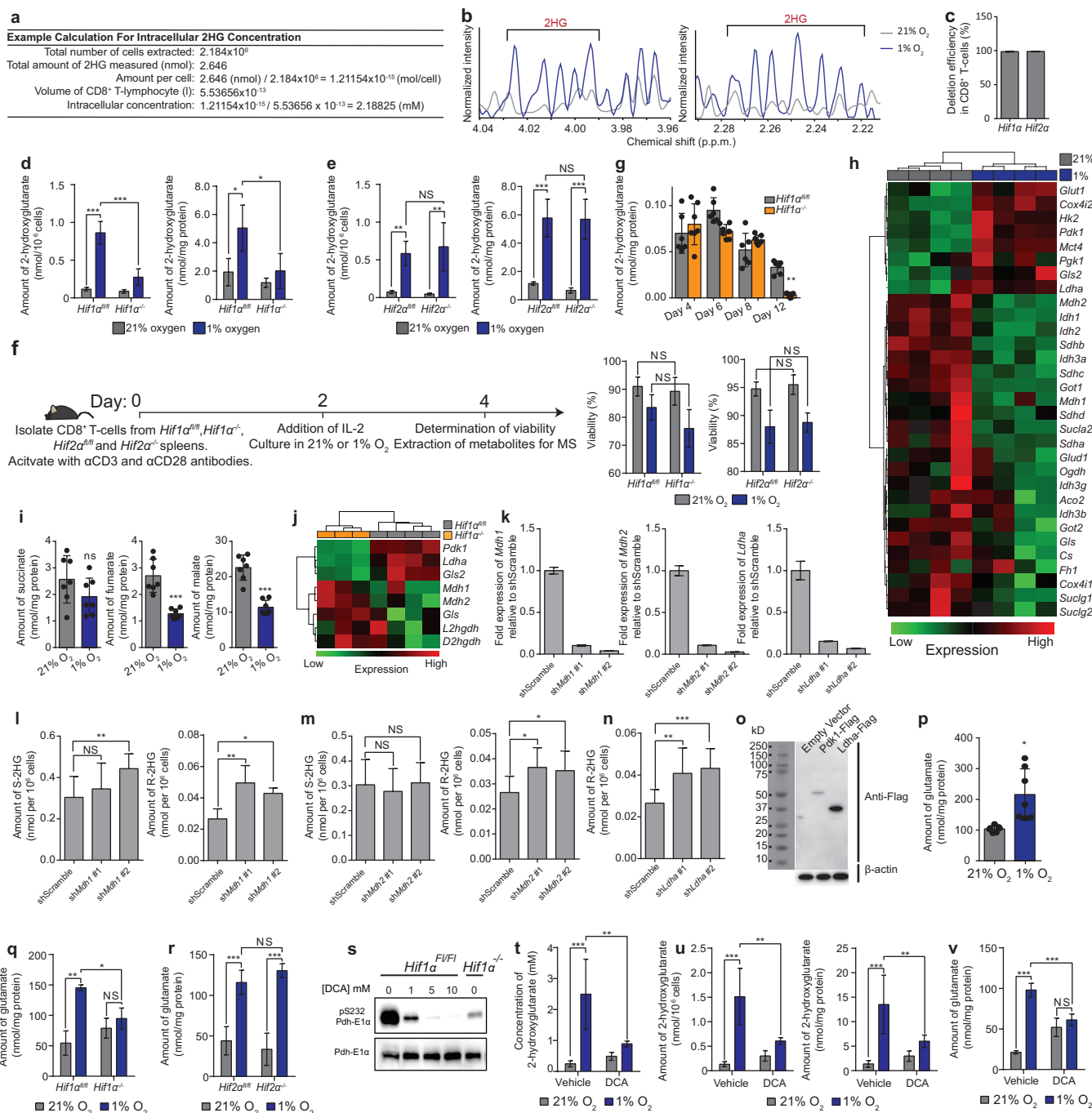
Author Contributions P.A.T. and A.P. designed experiments and carried out *ex vivo* and *in vivo* experiments, analysed the data, and wrote the manuscript. A.T.P., A.D., A.W.G. and R.S.J. carried out the initial metabolome survey. D.M., P.V., J.S., A.A. and C.E.E. aided in experimental execution and analysis. K.L.L., G.S.C., J.Y. and L.P. carried out ChIP and DIP experiments. J.R.G., K.L.L., L.P. and A.W.G. designed experiments and analysed data. R.S.J. designed experiments, analysed data, wrote the manuscript and administered the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.S.J. (rsj33@cam.ac.uk).



Extended Data Figure 1 | VHL-HIF-1 α regulate central carbon metabolism and 2HG levels in CD8⁺ T lymphocytes. **a**, Illustration of central carbon metabolism in CD8⁺ T lymphocytes, including glycolysis and the tricarboxylic acid cycle, depicting relative levels of detected metabolites between *Vhl*^{fl/fl} ($n=5$), *Vhl*^{fl/fl}*dLck*^{cre} ($n=5$) and *Hif-1 α* ^{fl/fl}*Vhl*^{fl/fl}*dLck*^{cre} ($n=3$) CD8⁺ T-lymphocyte groups. **b**, Glucose consumption and lactate production in *Vhl*^{fl/fl} and *Vhl*^{fl/fl}*dLck*^{cre} CD8⁺ T lymphocytes 7 days after activation with anti-CD3 and anti-CD28 antibodies ($n=4$ mice per genotype). **c**, Immunoblot analysis for Hif-1 α and LaminB1, using nuclear extracts prepared from *Vhl*^{fl/fl} and *Vhl*^{-/-} CD8⁺ T lymphocytes cultured

in 21% oxygen. **d**, Rank of metabolite loadings in PC1 from PCA. **e**, Immunoblot analysis for HIF-1 α , HIF-2 α and β -tubulin, on whole-cell extracts prepared from RCC4 and 786-O renal cancer cell lines, with and without expression of functional VHL. **f**, Deletion efficiency of *Vhl* in *Vhl*^{fl/fl} MEFs following infection with adeno-Cre virus, $n=3$ individual preparations. Accompanying immunoblot analysis for HIF-1 α and β -tubulin, on whole-cell extracts. Two-tailed *t*-test (**b**), one-way ANOVA for multiple comparisons (**a**). Error bars denote s.d. and each dot represents an individual mouse in **a** and **b**. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. For immunoblot source images, see Supplementary Fig. 1.

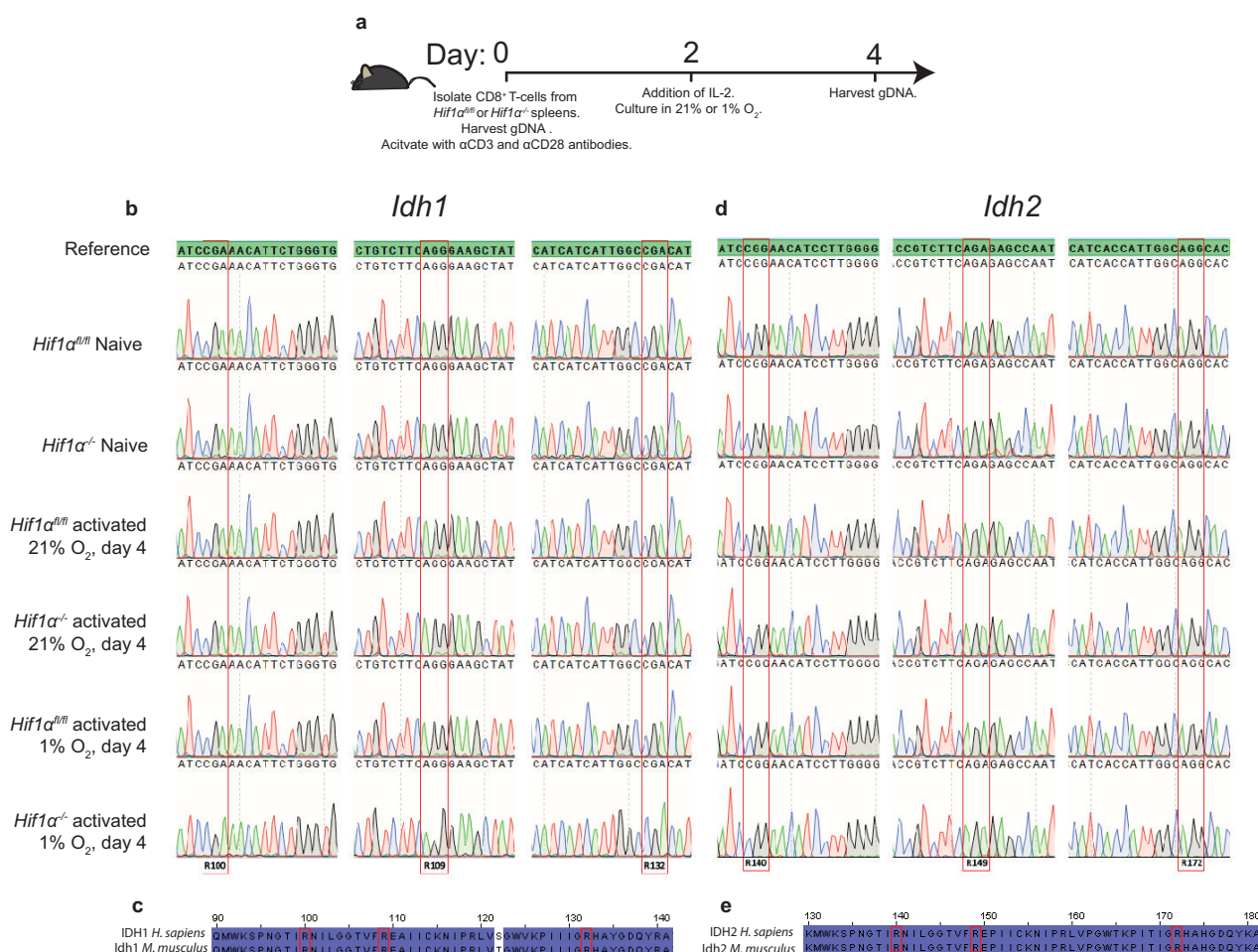


Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Hif-1 α -dependent metabolic alterations

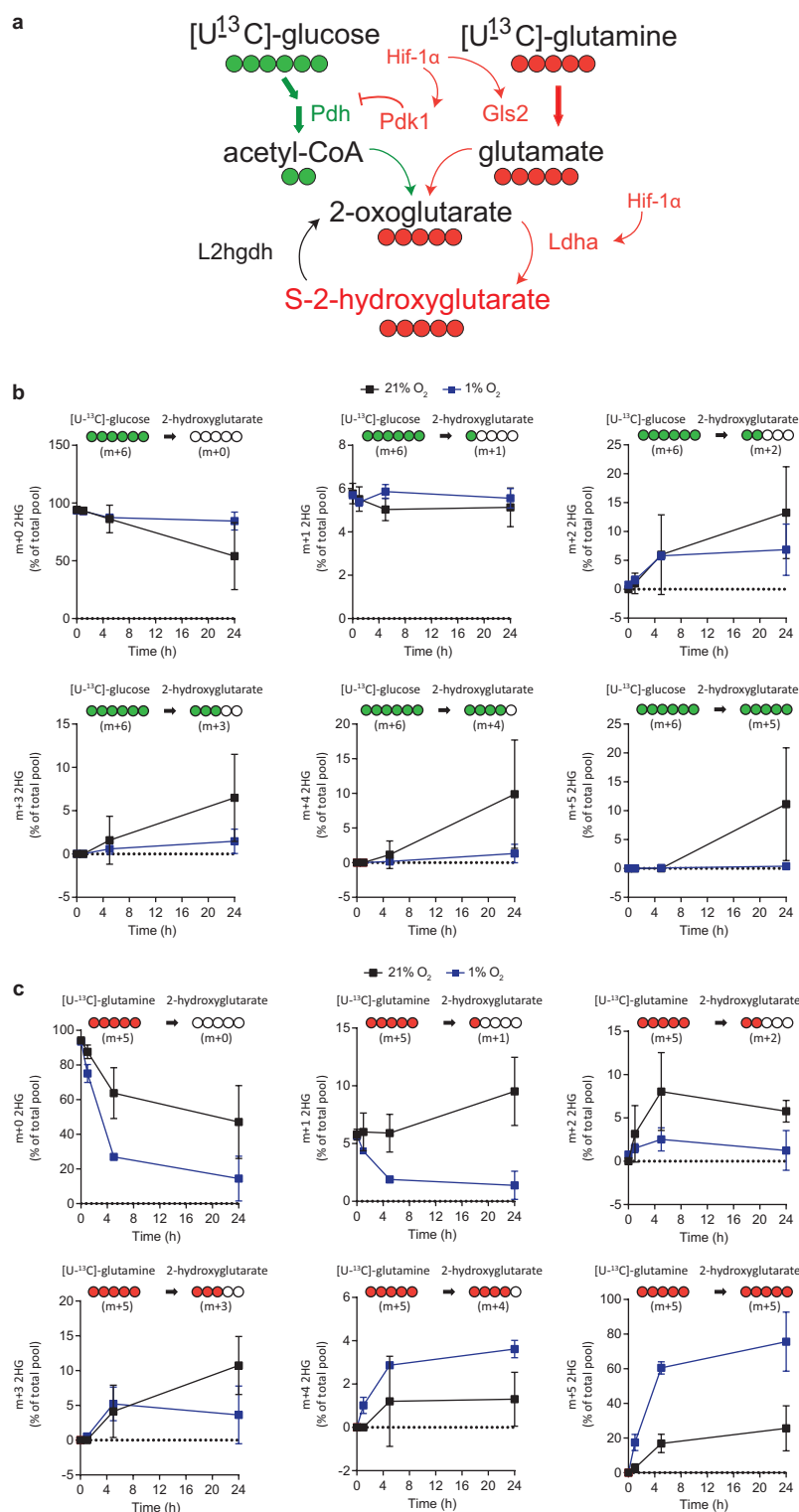
underlie S-2HG production in CD8⁺ T lymphocytes. **a**, Example calculation of intracellular 2HG concentration. **b**, ¹H-NMR analysis for 2HG from CD8⁺ T lymphocytes cultured as in Fig. 2b. **c**, Deletion efficiency of *Hif1a* or *Hif2a* in CD8⁺ T lymphocytes, isolated from *Hif1a*^{fl/fl} *dLck*^{cre} or *Hif2a*^{fl/fl} *dLck*^{cre} mice ($n = 4$ mice). **d, e**, Total 2HG levels, normalized to viable cell count or protein content, in *Hif1a*^{fl/fl} and *Hif1a*^{fl/fl} *dLck*^{cre} (**d**) and *Hif2a*^{fl/fl} and *Hif2a*^{fl/fl} *dLck*^{cre} (**e**) CD8⁺ T lymphocytes cultured as in Fig. 2b ($n = 4$ mice per genotype). **f**, Illustration outlining the workflow for metabolite extraction, deletion efficiency and viability experiments in *Hif1a*^{fl/fl}, *Hif1a*^{fl/fl} *dLck*^{cre}, *Hif2a*^{fl/fl} and *Hif2a*^{fl/fl} *dLck*^{cre} CD8⁺ T lymphocytes. Also shown are viability measurements at day 4 ($n = 4$ mice per genotype). **g**, Total amount of 2HG in *Hif1a*^{fl/fl} ($n = 6$) and *Hif1a*^{fl/fl} *dLck*^{cre} ($n = 7$) CD8⁺ T lymphocytes, at indicated times following activation ($n \geq 4$ mice per time point). **h**, Heat map indicating qPCR measurement of expression of enzymes involved in central carbon metabolism in CD8⁺ T lymphocytes cultured as in Fig. 2b ($n = 4$ mice per condition). **i**, Liquid chromatography–tandem mass spectrometry (LC–MS/MS) quantification of total intracellular succinate, fumarate and malate levels in CD8⁺ T lymphocytes isolated from C57BL/6J mice and cultured as in Fig. 2b ($n = 7$ mice). **j**, Heat map indicating qPCR measurement, in *Hif1a*^{fl/fl} ($n = 4$) and *Hif1a*^{fl/fl} *dLck*^{cre} ($n = 3$) CD8⁺ T lymphocytes growing in 1% oxygen, of expression of enzymes implicated in the hypoxic production of S-2HG. **k**, qPCR validation of

shRNA-knockdowns in CD8⁺ T lymphocytes isolated from C57BL/6J mice. **l–n**, LC–MS/MS quantification of S- and R-2HG in CD8⁺ T lymphocytes isolated from C57BL/6J mice, with shRNA-mediated knockdown of *Mdh1* (**l**), *Mdh2* (**m**) and *Ldha* (**n**) ($n = 4$ pools of 4 mice per pool). **o**, Validation of Pdk1–Flag and Ldha–Flag expression in *Hif1a*^{fl/fl} *dLck*^{cre} CD8⁺ T lymphocytes by immunoblot analysis for Flag. **p**, LC–MS/MS quantification of total intracellular glutamate levels in CD8⁺ T lymphocytes cultured as in Fig. 2b; $n = 7$ mice. **q, r**, LC–MS/MS quantification of total intracellular glutamate levels in *Hif1a*^{fl/fl} *dLck*^{cre}, *Hif2a*^{fl/fl} and *Hif2a*^{fl/fl} *dLck*^{cre} CD8⁺ T lymphocytes cultured as in Fig. 2b; $n = 4$ mice per genotype. **s**, Immunoblot of cytosolic fractions for phospho-Pdh-E1 α (S232) and total Pdh-E1 α in CD8⁺ T lymphocytes cultured in 1% oxygen in the presence of the indicated concentration of DCA for 48h. **t–v**, Total intracellular concentration of 2HG (**t**), 2HG normalized to viable cell count or protein content (**u**) and glutamate (**v**) in CD8⁺ T lymphocytes from C57BL/6J mice cultured as in Fig. 2b and treated with 5mM DCA for the latter 48 h of culture ($n = 4$ mice). Two-way ANOVA for grouped data (**d–f**, **q, r**, **t–v**). Paired *t*-test for matched comparisons (**i**, **p**), one-way ANOVA for multiple matched comparisons (**g**, **l–n**). Error bars denote s.d. and each dot represents an individual mouse in **g**, **i** and **p**. NS, non-significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$. Experiments were performed with indicated numbers of mice from multiple occasions. For immunoblot source images, see Supplementary Fig. 1.



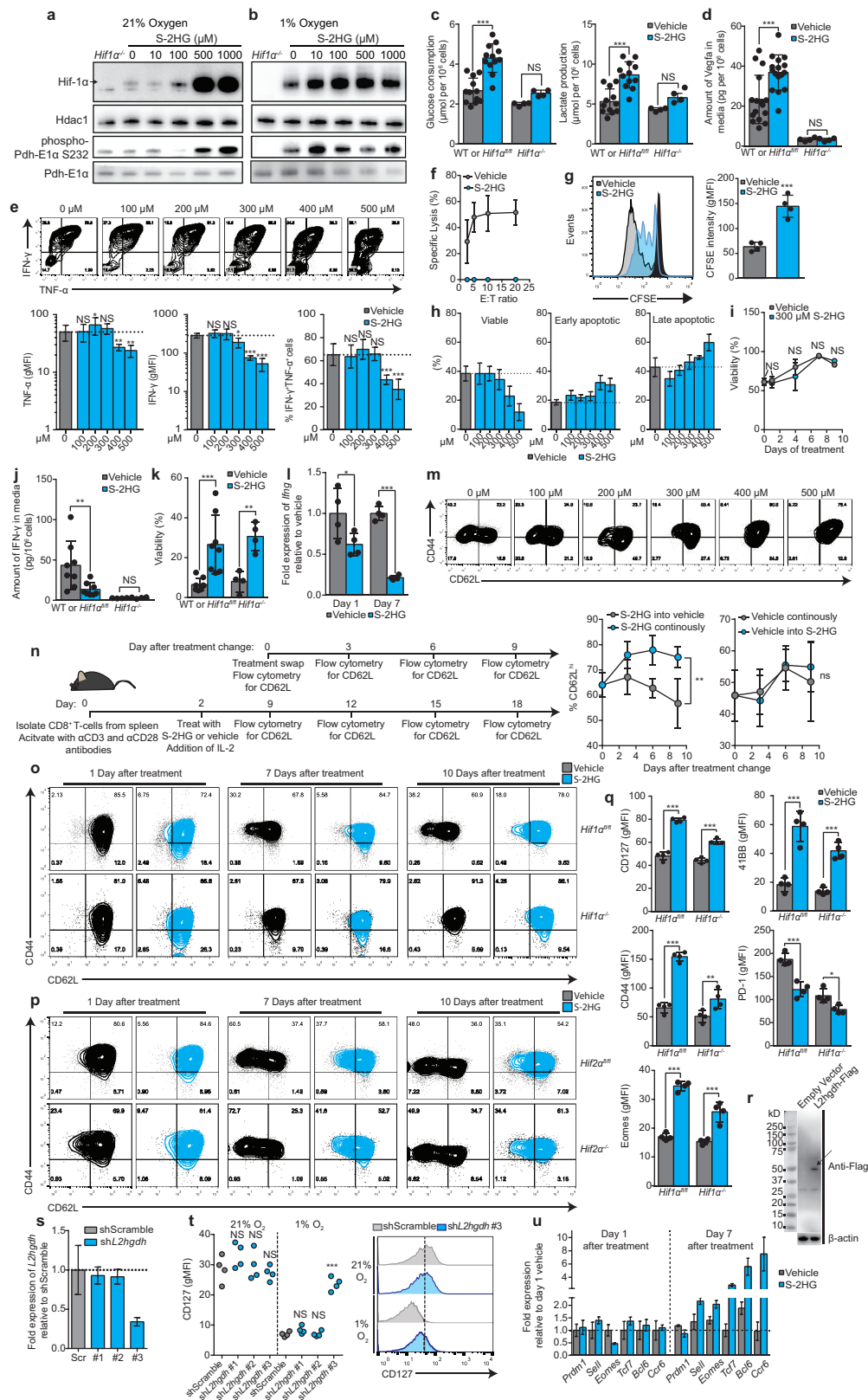
Extended Data Figure 3 | Naive and expanding primary CD8⁺ T lymphocytes do not possess mutations in *Idh1* or *Idh2* that can explain the presence of high levels of 2HG. **a**, Illustration outlining the workflow for mutational analysis of *Idh1* and *Idh2*. **b**, Sanger sequencing chromatograms validating the presence of wild-type *Idh1* as compared to the C57BL/6J NCBI reference sequence. **c**, Alignment of mouse and

human IDH1 protein indicating conservation of active site arginine residues. **d**, Sanger sequencing chromatograms validating the presence of wild-type *Idh2* as compared to the C57BL/6J NCBI reference sequence. **e**, Alignment of mouse and human IDH2 protein indicating conservation of active site arginine residues.



Extended Data Figure 4 | Kinetics of 2HG labelling in 21% and 1% oxygen, by $U\text{-}^{13}\text{C}$ -glucose and $U\text{-}^{13}\text{C}$ -glutamine. **a**, Proposed mechanism by which Hif-1 α controls S-2HG production in CD8^+ T lymphocytes and ^{13}C -labelling strategy using $U\text{-}^{13}\text{C}$ -glucose (m+6) and $U\text{-}^{13}\text{C}$ -glutamine (m+5) to label endogenous 2HG. Red and green represent pathways promoted and inhibited respectively by HIF-1 α in hypoxia. **b**, Isotopologue distribution of 2HG (as a percentage of the total pool)

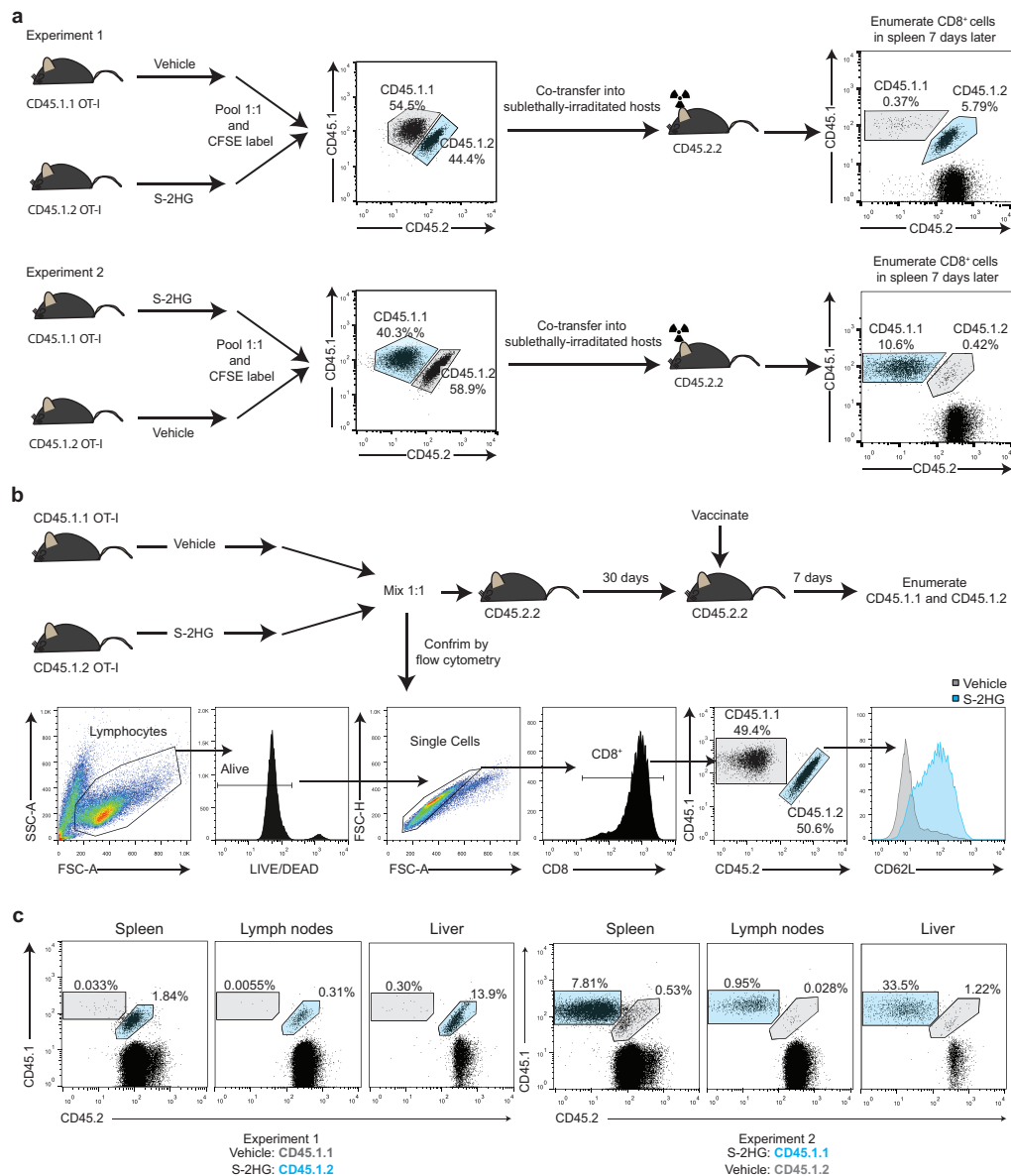
in CD8^+ T lymphocytes, after labelling with $U\text{-}^{13}\text{C}$ -glucose for 1, 5 and 24 h in both 21% and 1% oxygen conditions ($n = 3$ mice per time point). **c**, Isotopologue distribution of 2HG (as a percentage of the total pool) in CD8^+ T lymphocytes, after labelling with $U\text{-}^{13}\text{C}$ -glutamine for 1, 5 and 24 h in both 21% and 1% oxygen conditions ($n = 3$ mice per time point). Error bars, s.d.



Extended Data Figure 5 | See next page for caption.

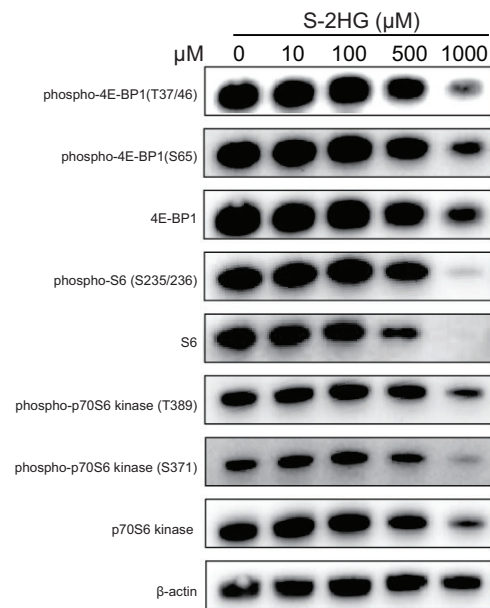
Extended Data Figure 5 | S-2HG treatment promotes Hif-1 α stability and alters the phenotypic and functional properties of CD8⁺ T lymphocytes in a Hif-1 α -independent manner. **a, b**, Immunoblot analysis of nuclear and cytosolic fractions, prepared from CD8⁺ T lymphocytes cultured in 21% (**a**) and 1% (**b**) oxygen, for Hif-1 α , Hdac1, phospho-Pdh-E1 α (S232) and total Pdh-E1 α . Cells were activated for 48 h with anti-CD3 and anti-CD28 antibodies and then expanded for a further 4 days in the presence of IL-2, followed by treatment with the indicated concentration of S-2HG for 16 h. The arrow indicates Hif-1 α protein. **c**, Glucose consumption and lactate production of C57BL/6J, *Hif1a*^{fl/fl} ($n = 12$) and *Hif1a*^{fl/fl}*dLck*^{cre} ($n = 4$) CD8⁺ T lymphocytes treated with or without 500 μ M S-2HG-octyl ester. **d**, Vegfa production of wild-type C57BL/6J, *Hif1a*^{fl/fl} ($n = 16$) and *Hif1a*^{fl/fl}*dLck*^{cre} ($n = 4$) CD8⁺ T lymphocytes treated with or without 500 μ M S-2HG-octyl ester. **e**, Representative flow cytometry plots of IFN- γ versus TNF- α in SIINFEKL re-stimulated OT-I CD8⁺ T lymphocytes, as a function of increasing doses of S-2HG-octyl ester for 7 days. Associated quantification and statistics are shown in the graphs below. **f**, Specific killing of EG7-OVA cells by OT-I CD8⁺ T lymphocytes ($n = 3$ mice per condition). **g**, CFSE dilution assay ($n = 4$ mice per condition) at day 3 of CD8⁺ T lymphocytes activated with anti-CD3 and anti-CD28 antibodies and cultured with or without 500 μ M S-2HG-octyl ester from day 0. Associated quantification and statistics are shown in the graph on the right. **h**, Viability and annexin V assay of CD8⁺ T lymphocytes treated with increasing S-2HG doses for 4 days ($n = 4$ mice). **i**, Viability of CD8⁺ T lymphocytes cultured with 300 μ M S-2HG-octyl ester for the indicated number of days ($n = 4$ mice). **j**, Amount of IFN- γ protein in the media of wild-type C57BL/6J, *Hif1a*^{fl/fl} ($n = 8$) and *Hif1a*^{fl/fl}*dLck*^{cre} ($n = 4$) CD8⁺ T lymphocytes treated for 24 h with or without 500 μ M S-2HG-octyl ester. **k**, Viability of *Hif1a*^{fl/fl} ($n = 8$) and *Hif1a*^{fl/fl}*dLck*^{cre} ($n = 4$) OT-I CD8⁺ T lymphocytes activated with 1,000 nM SIINFEKL peptide and cultured for 7 days with or without 500 μ M S-2HG-octyl ester in the absence of IL-2 supplementation from day 0. **l**, Expression of *Ifng* mRNA in CD8⁺ T lymphocytes treated for

either 24 h or 7 days with or without 500 μ M S-2HG-octyl ester. $n = 4$ mice per group. **m**, CD44 and CD62L surface expression on OT-I CD8⁺ T lymphocytes treated with increasing doses of S-2HG for 7 days. Cells were activated with 1,000 nM SIINFEKL peptide; $n = 3$ mice. Gated on live, CD8⁺ cells. **n**, Illustration outlining the workflow for the experiment. Percentage of CD62L^{hi} CD8⁺ T lymphocytes, treated for 7 days with 500 μ M S-2HG-octyl ester (left) or vehicle (right), followed by washout or maintenance of the compound and follow up every three days, for 9 more days ($n = 4$ mice). Gated on live, CD8⁺ cells. **o, p**, CD44 and CD62L surface expression on *Hif1a*^{fl/fl} and *Hif1a*^{fl/fl}*dLck*^{cre} (**o**) or *Hif2a*^{fl/fl} and *Hif2a*^{fl/fl}*dLck*^{cre} (**p**) CD8⁺ T lymphocytes treated with or without 500 μ M S-2HG-octyl ester for 1, 7 and 10 days following treatment. Data are representative of 3 (**o**) or 2 (**p**) mice per genotype. Gated on live, CD8⁺ cells. **q**, Flow cytometric characterization of indicated phenotypic markers on *Hif1a*^{fl/fl} and *Hif1a*^{fl/fl}*dLck*^{cre} ($n = 4$) CD8⁺ T lymphocytes treated for 7 days with 500 μ M S-2HG-octyl ester. Gated on live, CD8⁺ cells. **r**, Validation of L2hgdh-Flag expression in CD8⁺ T lymphocytes from C57BL/6J mice by immunoblot analysis for Flag. The arrow indicates L2hgdh-Flag protein. **s**, qPCR validation of *L2hgdh* knockdown in CD8⁺ T lymphocytes isolated from C57BL/6J mice. **t**, CD127 surface expression in response to *L2hgdh* knockdown ($n = 4$). Representative flow cytometry histogram of CD127 surface levels on transduced (GFP⁺) CD8⁺ T lymphocytes in response to shScramble or sh*L2hgdh* 3 in 21% or 1% oxygen is shown on the right. **u**, qPCR quantification of *Prdm1*, *Sell*, *Eomes*, *Tcf7*, *Bcl6* and *Ccr6* expression in CD8⁺ T lymphocytes treated for 1 or 7 days with or without 500 μ M S-2HG-octyl ester. Paired *t*-test for matched comparisons (**g**) and two-way ANOVA for grouped data (**c, d, j, k, l, q**). One-way ANOVA for multiple comparisons (**i, n, t**). Error bars denote s.d. and each dot in **c, d, g, j, k, l, q, t** represents an individual mouse. NS, non-significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. gMFI, geometric mean fluorescence intensity. Experiments were performed with indicated numbers of mice from multiple occasions. For immunoblot source images, see Supplementary Fig. 1.

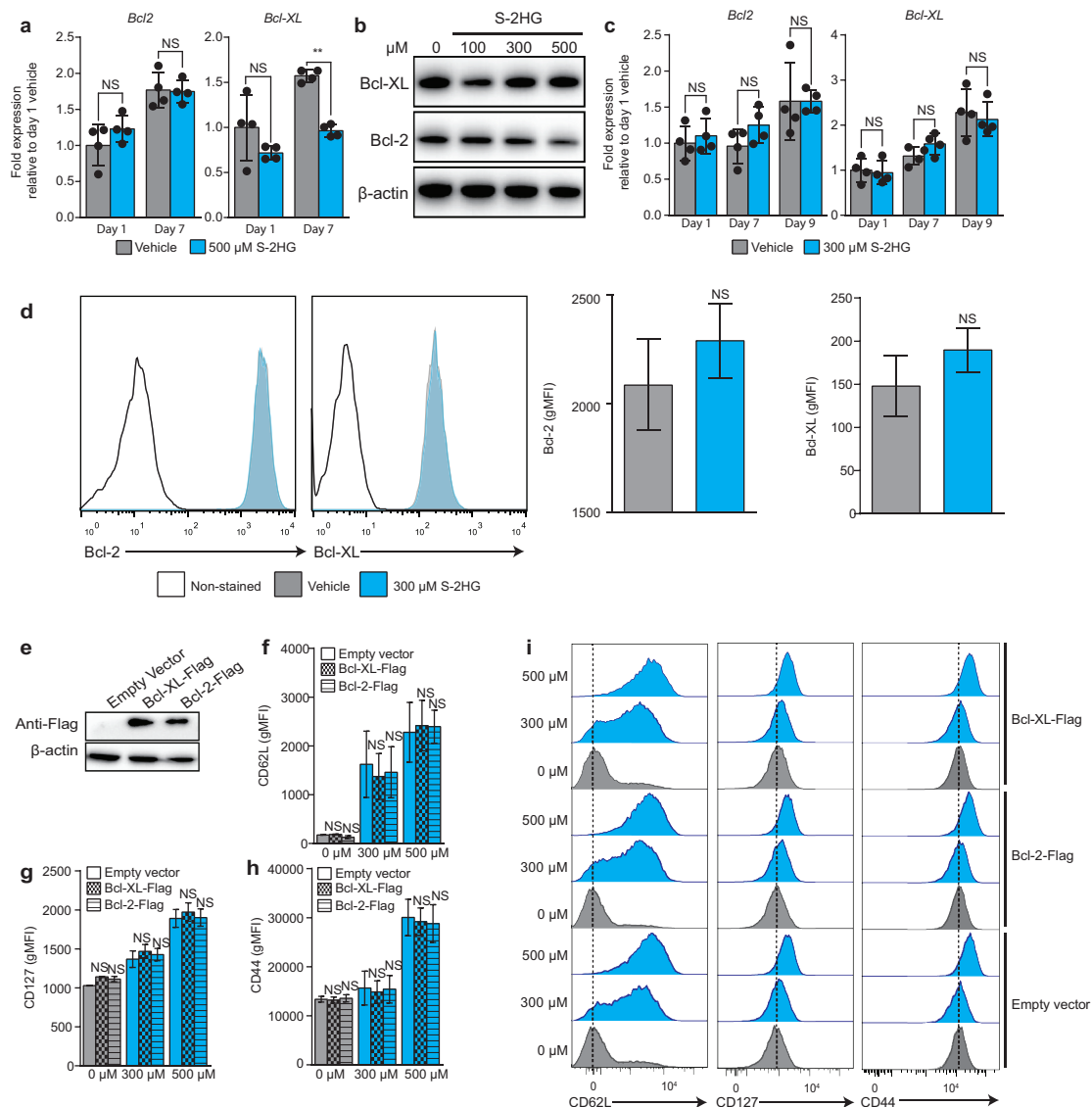


Extended Data Figure 6 | Ex vivo treatment of CD8⁺ T lymphocytes with S-2HG promotes *in vivo* homeostatic proliferation and recall of adoptively transferred cells. **a**, Diagram outlining the homeostatic proliferation experiments in Fig. 4a–c. Representative flow cytometry plots are shown for each pool before and after adoptive transfer. Flow

cytometry plots show viable CD8⁺ cells. **b**, Diagram outlining the recall experiments in Fig. 4f. **c**, Representative flow cytometry plots of recalling CD45.1⁺CD8⁺ T lymphocytes in indicated organs on day 7 after vaccination (day 37 after transfer).

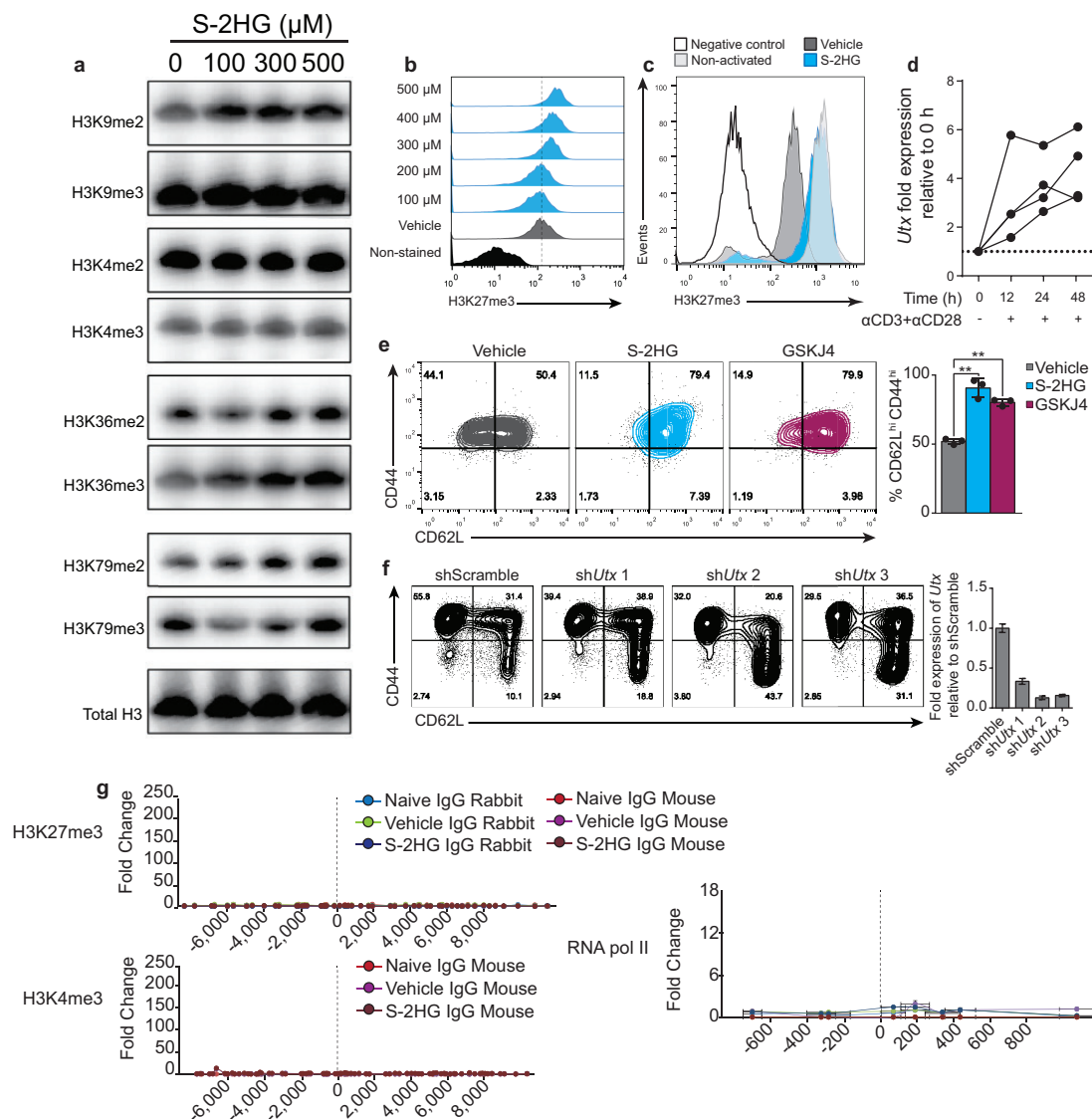


Extended Data Figure 7 | S-2HG does not inhibit mTOR signalling at the doses necessary for the formation of memory-like CD8⁺ T lymphocytes. Immunoblot analysis on cytosolic extracts for mTOR signalling in CD8⁺ T lymphocytes treated with the indicated doses of S-2HG for 24 h. For immunoblot source images, see Supplementary Fig. 1.



Extended Data Figure 8 | S-2HG does not induce Bcl-2 or Bcl-XL that can explain the *in vivo* persistence of adoptively transferred CD8⁺ T lymphocytes. **a**, qPCR quantification of *Bcl2* and *Bcl2l1* (*Bcl-XL*) mRNA levels in response to 500 μ M S-2HG-octyl ester treatment for either 1 or 7 days ($n = 4$ mice). **b**, Immunoblot analysis for Bcl-2 and Bcl-XL protein in response to increasing doses of S-2HG-octyl ester for 9 days. **c**, qPCR quantification of *Bcl2* and *Bcl-XL* mRNA levels in response to 300 μ M S-2HG-octyl ester treatment for either 1, 7 or 9 days ($n = 4$ mice). **d**, Representative flow cytometry histograms of Bcl-2 and Bcl-XL abundance in CD8⁺ T lymphocytes treated with 300 μ M S-2HG-octyl ester for 9 days. Quantification and associated statistics are shown in the graph on the right ($n = 3$ mice). **e**, Immunoblot analysis confirming the expression of Bcl-XL-Flag and Bcl-2-Flag in OT-I in CD8⁺ T lymphocytes. **f-h**, CD62L (**f**), CD127 (**g**) and CD44 (**h**) surface expression in OT-I

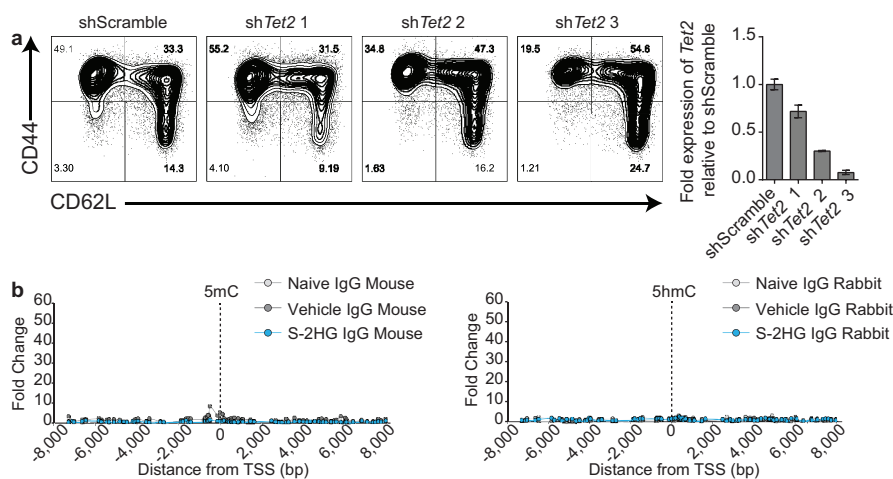
CD8⁺ T lymphocytes transduced with retrovirus expressing either Bcl-2-Flag or Bcl-XL-Flag and treated with the indicated concentration of S-2HG-octyl ester for 7 days ($n = 2$ mice). **i**, Representative flow cytometry histograms of CD62L, CD127 and CD44 surface expression in OT-I CD8⁺ T lymphocytes transduced with retrovirus expressing either Bcl-2-Flag or Bcl-XL-Flag and treated with the indicated concentration of S-2HG-octyl ester for 7 days. The associated statistics of these flow cytometry data are shown in **f**, **g** and **h**. ****** $P < 0.01$, NS, non-significant. Paired *t*-test for matched comparisons (**d**) and two-way ANOVA for grouped data (**a**). One-way ANOVA of matched samples for multiple comparisons (**c**, **f**, **g**, **h**). Error bars denote s.d. and each dot in **a** and **c** represents an individual mouse. Experiments were performed with indicated numbers of mice from at least two occasions. For immunoblot source images, see Supplementary Fig. 1.



Extended Data Figure 9 | S-2HG induces global histone H3 methylation changes in CD8⁺ T lymphocytes. **a**, Immunoblot analysis on nuclear extracts for histone H3 methylation marks in activated CD8⁺ T lymphocytes treated with the indicated doses of S-2HG for 7 days.

b, Representative flow cytometry histograms of H3K27me3 staining as a function of increasing S-2HG-octyl ester concentration. **c**, H3K27me3 staining in CD8⁺ T lymphocytes treated with or without 500 μM S-2HG-octyl ester and stained with or without fluorophore-conjugated C36B11 antibody. **d**, qPCR measurement for expression of *Utx* in unstimulated and stimulated CD8⁺ T lymphocytes; $n = 4$ mice. Expression for *Utx* is displayed for each mouse individually. **e**, Representative flow cytometry plots of CD44 versus CD62L expression, with associated statistics, on activated CD8⁺ T lymphocytes after 4 days of treatment with 500 μM S-2HG-octyl ester or 1 μM GSKJ4. Gated on live, CD8⁺ cells. $n = 3$ mice.

f, Representative flow cytometry plots of CD44 versus CD62L expression on CD8⁺ T lymphocytes with shRNA-mediated knockdown of *Utx*, 7 days after transduction. Gated on live CD8⁺GFP⁺ cells. Graph on right shows knockdown hairpin fidelity for *Utx*. **g**, IgG control ChIP-qPCR for H3K4me3, H3K27me and RNA Pol II at and around the TSS for CD62L, in freshly isolated naive or activated CD8⁺ T lymphocytes treated with or without 500 μM S-2HG-octyl ester for 7 days. Each profile shows the fold change over the non-binding control primer. Each dot represents an individual primer pair. A pool of $n = 6$ mice was used for each condition and error bars denote s.e.m. One-way ANOVA for multiple matched comparisons (**e**). Each dot in **e** represents an individual mouse. Error bars (**e**, **f**) denote s.d. $**P < 0.01$. Experiments were performed with indicated numbers of mice from at least two occasions. For immunoblot source images, see Supplementary Fig. 1.



Extended Data Figure 10 | S-2HG induces global changes in the content of 5hmC and 5mC in genomic DNA of CD8⁺ T-lymphocyte genomic DNA. **a**, Representative flow cytometry plots of CD44 versus CD62L expression on CD8⁺ T lymphocytes with shRNA-mediated knockdown of *Tet2*, 7 days after transduction. Gated on live CD8⁺GFP⁺ cells. Graph on right shows knockdown hairpin fidelity for *Tet2* and error bars denote s.d.

b, IgG control DIP-qPCR for 5mC and 5hmC at and around the TSS for CD62L, in freshly isolated naive or activated CD8⁺ T lymphocytes treated with or without 500 μ M S-2HG for 7 days. Each profile shows the fold change over the non-binding control primer. Each dot represents an individual primer pair. A pool of $n = 6$ mice was used for each condition and error bars denote s.e.m.

m⁶A modulates neuronal functions and sex determination in *Drosophila*

Tina Lence¹, Junaid Akhtar¹, Marc Bayer¹, Katharina Schmid², Laura Spindler³, Cheuk Hei Ho⁴, Nastasja Kreim¹, Miguel A. Andrade-Navarro^{1,5}, Burkhard Poeck³, Mark Helm² & Jean-Yves Roignant¹

N⁶-methyladenosine RNA (m⁶A) is a prevalent messenger RNA modification in vertebrates. Although its functions in the regulation of post-transcriptional gene expression are beginning to be unveiled, the precise roles of m⁶A during development of complex organisms remain unclear. Here we carry out a comprehensive molecular and physiological characterization of the individual components of the methyltransferase complex, as well as of the YTH domain-containing nuclear reader protein in *Drosophila melanogaster*. We identify the member of the split ends protein family, Spenito, as a novel bona fide subunit of the methyltransferase complex. We further demonstrate important roles of this complex in neuronal functions and sex determination, and implicate the nuclear YT521-B protein as a main m⁶A effector in these processes. Altogether, our work substantially extends our knowledge of m⁶A biology, demonstrating the crucial functions of this modification in fundamental processes within the context of the whole animal.

RNA modifications represent a critical layer of epigenetic regulation of gene expression¹. m⁶A is among the most abundant modifications in the mammalian system^{2,3}. m⁶A distribution has been determined in several organisms and cell types, including human, mouse, rice and yeast^{4–7}. The modification is found in a subset of the RRACH consensus sites (R, purine; H, non-guanine base) and is enriched around stop codons, in the 3′-untranslated regions (3′ UTRs) and within long internal exons. m⁶A was shown to control several post-transcriptional processes, including pre-mRNA splicing, mRNA decay and translation^{4,5,8–16}, which are mediated in part via conserved members of the YTH protein family^{4,17}. The methyltransferase complex catalysing m⁶A formation in mammals consists of methyltransferase-like 3 (METTL3), methyltransferase-like 14 (METTL14) and a stabilizing factor called Wilms' tumour 1-associated protein (WTAP)^{9,11,18,19}. In mammals, m⁶A can be reverted into adenosine via two identified demethylases: fat mass and obesity associated factor (FTO)^{20–22} and AlkB homologue 5 (ALKBH5)²³.

Several studies have uncovered crucial roles for METTL3 during development and cell differentiation. Knockout of *Mettl3* in murine naive embryonic stem cells blocks differentiation^{24,25}, while its deletion in mice causes early embryonic lethality. Similarly, in *Drosophila*, loss of the *METTL3* orthologue *Ime4* is reported to be semi-lethal during development, with adult escapers having reduced fertility owing to impaired Notch signalling²⁶. Depletion of the *METTL3* orthologue *MTA* in *Arabidopsis thaliana* also affects embryonic development^{27,28}, while in yeast *ime4* has an essential role during meiosis^{29–31}. All of these observations indicate the importance of m⁶A in the gonads and during early embryogenesis. Recent crystal structure studies investigated the molecular activities of the two predicted catalytic proteins^{32,33}, however, their respective roles *in vivo* remain unclear. Here we characterize members of the methyltransferase complex in *Drosophila* and identify the split ends (SPEN) family protein, Spenito (Nito), as a novel bona fide subunit. Expression of complex components is substantially enriched in the nervous system, and flies with mutations in *Ime4* and *Mettl14* suffer from impaired neuronal functions. Methyltransferase complex components also influence the female-specific splicing of

Sex-lethal (*Sxl*), revealing a role in fine-tuning sex determination and dosage compensation. Notably, knockout of the nuclear m⁶A reader YT521-B resembles the loss of the catalytic subunits, implicating this protein as a main effector of m⁶A *in vivo*.

m⁶A is enriched in the nervous system

To investigate potential functions of m⁶A in *Drosophila*, we monitored its levels on mRNA samples isolated at different developmental stages of wild-type flies using mass spectrometry (Fig. 1a and Extended Data Fig. 1a, b). We find that m⁶A is remarkably enriched in early embryogenesis but drops dramatically 2 h after fertilization and remains low throughout the rest of embryogenesis and early larval stages. During the third larval instar, m⁶A rises again to reach a peak at pupal phases. While the overall level of m⁶A decreases in adults, it remains substantially elevated in heads and ovaries.

A phylogenetic analysis of the *Drosophila* METTL3 orthologue Ime4 (Extended Data Fig. 1c) identifies two closely related factors, CG7818 and CG14906. Depletion of Ime4 and CG7818 in embryonic-derived Schneider (S2R+) cells decreases m⁶A levels by about 70%, whereas depletion of CG14906 had no effect (Fig. 1b and Extended Data Fig. 1d). These results indicate that Ime4 and CG7818 are required to promote m⁶A activity in *Drosophila*. Because of its sequence and functional conservation with human METTL14, CG7818 was renamed dMettl14. Fl(2)d and Virilizer (Vir) are the *Drosophila* homologues of WTAP and KIAA1429, respectively, which are integral components of the complex in mammals⁶. Both transcripts follow the same developmental distribution as other methyltransferase complex components and their depletion also affects m⁶A levels (Fig. 1a, b and Extended Data Fig. 1d–f). Ime4 and Fl(2)d co-immunoprecipitate with dMettl14 in an RNA-independent manner (Fig. 1c). Likewise, Vir, Fl(2)d and Ime4 are found in the same complex (Extended Data Fig. 1g). Notably, Fl(2)d depletion reduces the interaction between Ime4 and dMettl14 (Extended Data Fig. 1h), confirming its proposed role as a stabilizing factor. All components localize in the nucleus and are ubiquitously expressed in early embryonic stages (data not shown) but show substantial enrichment in the neuroectoderm at later stages (Fig. 1d, e).

¹Institute of Molecular Biology (IMB), 55128 Mainz, Germany. ²Institute of Pharmacy and Biochemistry, Johannes Gutenberg University of Mainz, 55128 Mainz, Germany. ³Institute of Zoology III (Neurobiology), Johannes Gutenberg University of Mainz, 55128 Mainz, Germany. ⁴Kimmel Center for Biology and Medicine of the Skirball Institute, NYU School of Medicine, Department of Cell Biology, 540 First Avenue, New York, New York 10016, USA. ⁵Faculty of Biology, Johannes Gutenberg University of Mainz, 55128 Mainz, Germany.

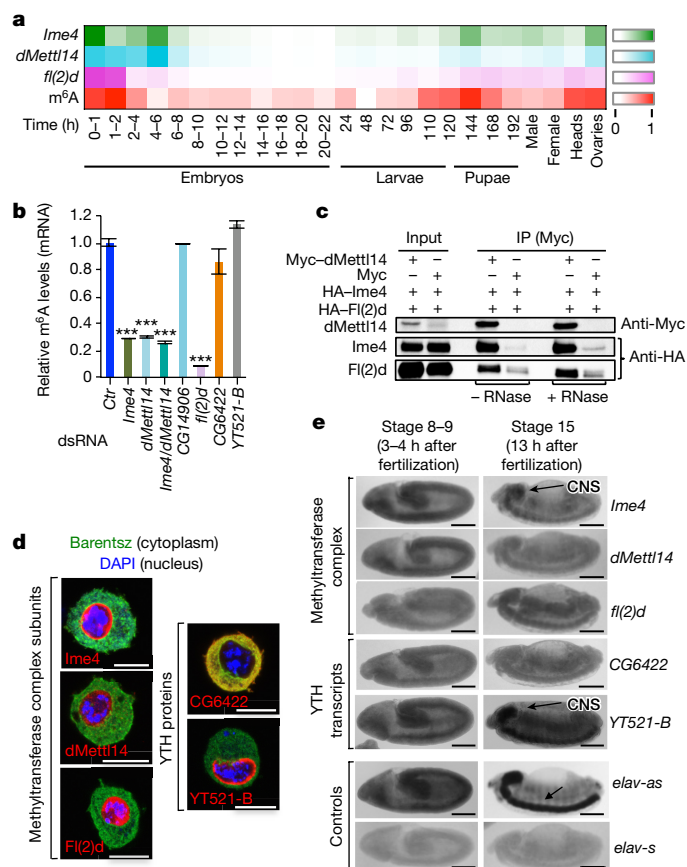


Figure 1 | *Drosophila* m⁶A methyltransferase complex is enriched in the nervous system. **a**, Heatmap shows relative mRNA expression of methyltransferase complex subunits and m⁶A levels during development. **b**, m⁶A liquid chromatography tandem-mass spectrometry (LC-MS/MS) quantification in different knockdown conditions. *Ctrl*, control. Bar chart represents the mean \pm standard deviation (s.d.) of three technical measurements from three biological replicates. *** $P < 0.0001$ (one-way analysis of variance (ANOVA), Tukey's post-hoc analysis). **c**, Lysates from S2R+ cells expressing indicated proteins were immunoprecipitated using Myc beads. HA, haemagglutinin. **d**, Immunostaining of indicated proteins in S2R+ cells. DAPI, 4',6-diamidino-2-phenylindole. Scale bars, 10 μ m. **e**, *In situ* RNA hybridization. *elav*-positive (*elav-as*) and -negative (*elav-s*) controls are shown. CNS, central nervous system. Scale bars, 100 μ m.

Altogether, our results demonstrate the existence of a conserved functional methyltransferase complex in *Drosophila* and reveal its particular abundance in the nervous system.

YT521-B mediates m⁶A-dependent splicing

To obtain insight into the transcriptome-wide m⁶A distribution in S2R+ cells, we performed methylated RNA immunoprecipitation followed by sequencing (MeRIP-seq) (Extended Data Fig. 2a, b). In total, 1,120 peaks representing transcripts of 812 genes were identified (Supplementary Table 1). The consensus sequence RRACH is present in most m⁶A peaks ($n = 1,027$, 92% of all peaks) (Fig. 2a). Additional sequences are also enriched, suggesting their potential involvement in providing specificity to the methyltransferase complex (Extended Data Fig. 2c). As shown in other species, enrichment near start and stop codons was observed (Fig. 2b, c).

We next performed transcriptome analyses in S2R+ cells lacking m⁶A components (Extended Data Fig. 3a–c). Knockdown of Fl(2)d leads to strong changes in gene expression ($n = 2,129$ differentially expressed genes; adjusted P value < 0.05 ; Fig. 2d, e and Supplementary Table 2), while knockdowns of Ime4 and dMettl14 have milder effects. Gene ontology analyses revealed that genes involved in diverse

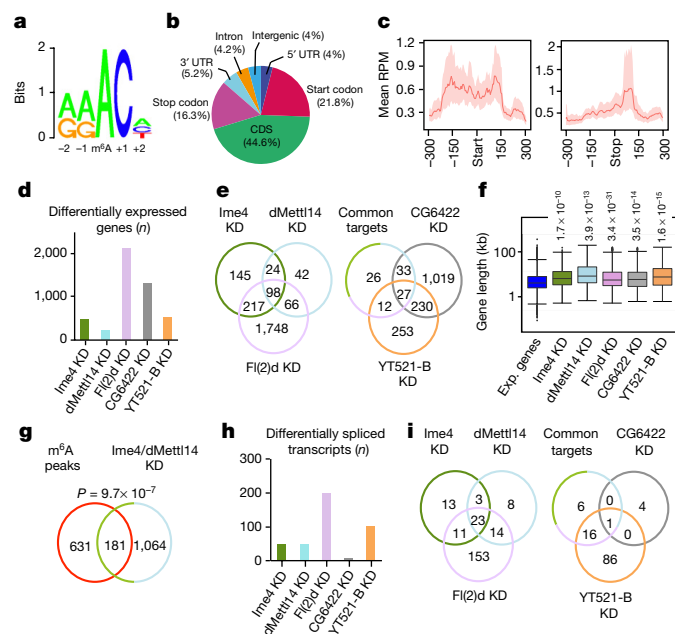


Figure 2 | m⁶A controls alternative splicing via YT521-B. **a**, Sequence logo of deduced consensus motif for most m⁶A peaks centred on the modified adenosine. **b**, Pie chart of m⁶A peak distribution in distinct transcript segments. Start codon (± 300 bp window around start), CDS (coding sequence (CDS) excluding 300 bp after start and 300 bp before stop), stop codon (± 300 bp window around stop). **c**, Normalized m⁶A read counts in a ± 300 bp window around start (left) and stop codons (right) for all transcripts detected. RPM, reads per million. **d**, Number of differentially expressed genes upon knockdown (KD) of indicated proteins. **e**, Venn diagrams representing common targets between methyltransferase complex components and YTH proteins. **f**, Boxplots of gene length for all expressed (Exp.) genes (average coverage > 1 read per kilobase per million mapped reads (RPKM) in control conditions) and differentially expressed genes in each knockdown. Distributions were compared to all expressed genes using the Wilcoxon rank sum test. Expressed genes were downsampled to the same number of genes as in the given knockdown. **g**, Venn diagram showing the overlap of differentially regulated and/or spliced genes in Ime4/dMettl14 double knockdown with genes containing m⁶A peaks. **h**, Number of differentially spliced genes upon knockdown of methyltransferase components and YTH proteins. **i**, Venn diagrams showing number of common targets between methyltransferase complex components and YTH proteins.

metabolic processes, anion transport and cell adhesion are significantly overrepresented (Extended Data Fig. 4). Despite the fact that S2R+ cells are of non-neuronal origin, the affected genes are also enriched for neuronal functions, including roles in axon guidance and synapse activity. Consistent with the larger average size of neuronal genes³⁴, affected genes are significantly larger than the non-affected ones (Fig. 2f). We next compared the genes affected upon Ime4/dMettl14 double knockdown with the m⁶A profile. Overall, about 15% of the affected genes contain at least one m⁶A peak (Fig. 2g). We find a slight but significant positive influence of m⁶A on mRNA levels (P value = 9.9×10^{-4}) (Extended Data Fig. 3d) and this effect seems independent of the location of the m⁶A peak along the transcript (Extended Data Fig. 3e). Several splicing changes upon knockdown of individual complex components were also observed (Fig. 2h, i). *fl(2)d* itself is among the affected transcripts in any of the knockdowns tested (Extended Data Fig. 5). Generally, each knockdown results in alternative 5' splice site usage and intron retention (Extended Data Fig. 3f), which was also observed in human cells⁴.

YTH proteins are critical readers of m⁶A in mammals. While vertebrates contain five proteins of this family, only two members exist in flies, CG6422 and YT521-B (Extended Data Fig. 6a). We find that

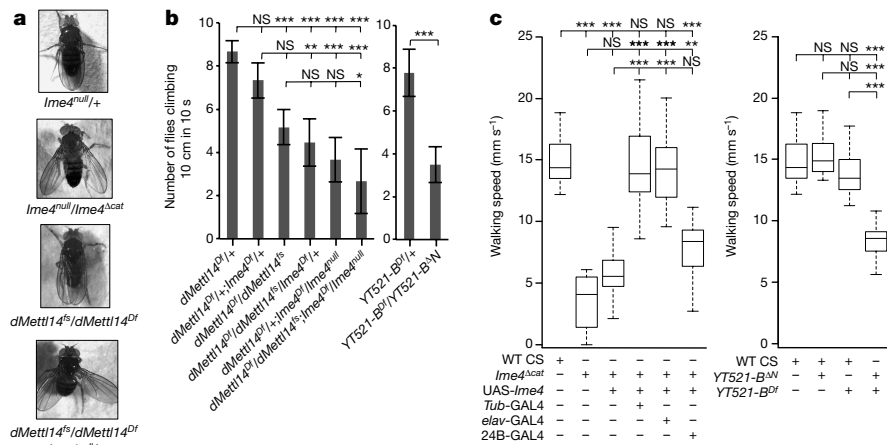


Figure 3 | YT521-B and the methyltransferase complex control fly behaviour. a, Adult flies of indicated genotypes. **b**, Bars represent the mean \pm s.d. of female flies ($n = 10$ per condition) that climb over 10 cm in 10 s (six independent measurements). * $P < 0.01$; ** $P < 0.001$; *** $P < 0.0001$; NS, not significant (one-way ANOVA, Tukey's post-hoc analysis). Student's t -test was used for YT521-B analysis. **c**, Walking

speed of indicated females ($n = 15$ per condition) measured in Buridan's paradigm. Kruskal–Wallis analysis with Bonferroni correction (*Ime4*) and one-way ANOVA, Bonferroni post-hoc analysis (YT521-B) were used. Boxes signify 25%/75% quartiles, thick lines indicate medians, and whiskers show maximum interquartile range $\times 1.5$. * $P < 0.0$, ** $P < 0.01$, *** $P < 0.001$; NS, not significant; WT-CS, wild-type *Canton-S* flies.

CG6422 is localized in the cytoplasm and strongly enriched during the first 2 h after fertilization but then declines and remains at low levels during development and adulthood. By contrast, YT521-B is strictly nuclear and shows strong enrichment in the embryonic nervous system and adult brains (Fig. 1d, e and Extended Data Fig. 6b). Using dot-blot assays and pull-down experiments we confirmed that YT521-B binds m⁶A in *Drosophila* (Extended Data Fig. 6c, d). RNA-sequencing (RNA-seq) experiments show that depletion of CG6422 only marginally affects splicing (Fig. 2h) while YT521-B knockdown significantly impairs this process (103 differentially regulated splicing events; adjusted P value < 0.1). The overlap of mis-spliced events between YT521-B knockdown and knockdown of methyltransferase complex subunits is about 70% (Fig. 2i), revealing that YT521-B might be the main mediator of m⁶A function in pre-mRNA splicing.

m⁶A components modulate fly behaviour

To investigate potential roles of m⁶A during *Drosophila* development, we generated *Ime4*- and *dMettl14*-knockout flies (Extended Data Fig. 7a, b, d). Two deletions in *Ime4* were created, removing the entire coding sequence (*Ime4^{null}*) or only the C-terminal part containing the catalytic domain (*Ime4^{cat}*). We find that flies homozygous for either mutant allele as well as transheterozygous flies survive until adulthood. We did not observe encapsulation defects in ovaries as previously shown using different alleles²⁶ (Extended Data Fig. 8a). However, the mutant flies have a reduced lifespan and exhibit multiple behavioural defects: flight and locomotion are severely affected and they spend more time grooming (Fig. 3b, c and Extended Data Fig. 8b). They also display a mild held-out wing appearance resulting from failure to fold their wings together over the dorsal surface of the thorax and abdomen (Fig. 3a). *dMettl14* mutant flies have normal wings but their locomotion is also deficient (Fig. 3a, b). To test whether *Ime4* and *dMettl14* can compensate for each other *in vivo*, we generated double-mutant animals. Removing one copy of *Ime4* in the *dMettl14* mutant background mimics the held-out wing phenotype observed upon loss of *Ime4* (Fig. 3a). Double-homozygous mutants give similar phenotypes as the *Ime4* single knockout, albeit with increased severity (Fig. 3b). Altogether, these phenotypic analyses strongly suggest that *Ime4* and *dMettl14* control similar physiological processes *in vivo*, indicating that they probably regulate common targets. Furthermore, the function of *Ime4* appears to be slightly predominant over *dMettl14* and most activities require its catalytic domain.

To quantify the locomotion phenotype better, the so-called Buridan's paradigm^{35,36} was applied. We find that the activity and walking speed of *Ime4* mutant flies is reduced by twofold compared with control flies (Fig. 3c and Extended Data Fig. 8c). In addition, orientation defects were observed. All phenotypes were rescued by ubiquitous (*Tub-GAL4*) and neuronal (*elav-GAL4*), but not mesodermal (*24B10-GAL4*) expression of *Ime4* complementary DNA (Fig. 3c and Extended Data Fig. 8c). These findings demonstrate that m⁶A controls *Drosophila* behaviour by specifically influencing neuronal functions. To investigate potential neurological defects underlying the behavioural phenotype, we examined the neuromuscular junction (NMJ) of *Ime4*-mutant larvae. Notably, NMJ synapses grow exuberantly in the *Ime4* mutant, displaying a 1.5-fold increase in the number of boutons and a 1.3-fold increase of active zones per bouton (Extended Data Fig. 7e), indicating that *Ime4* may regulate locomotion via control of synaptic growth at the NMJ. To identify target genes involved in locomotion, adult heads of 1–2-day-old female flies were dissected and subjected to RNA-seq. In total, 1,681 genes display significant changes in expression and splicing upon *Ime4* loss of function (adjusted P value < 0.05). Notably, many of the affected genes control fly locomotion ($n = 193$; Supplementary Table 6). We next compared the list of affected genes with our MeRIP data from S2R+ cells and identified a dozen locomotion-related genes as potential direct targets of m⁶A (Supplementary Table 7). Hence, it is likely that more than a single gene accounts for the locomotion phenotype observed in the absence of a functional methyltransferase complex.

m⁶A components modulate splicing of *Sxl* transcripts

Among the top hits showing changes in alternative splicing upon *Ime4* knockout was *Sxl*, encoding a master regulator of sex determination and dosage compensation³⁷ (Fig. 4a). *Sxl* is expressed in both females and males, but the transcript in males contains an additional internal exon introducing a premature stop codon. To confirm the role of *Ime4* and potentially *dMettl14* in *Sxl* splicing, we analysed RNA extracts from the heads of both sexes by polymerase chain reaction with reverse transcription (RT-PCR). While splicing is unaffected in males, mutant females of both genotypes show inclusion of the male-specific exon and decrease of the female-specific isoform (Fig. 4b, c). We found that this decrease is less pronounced when analysing isoform levels from whole flies, possibly reflecting the specific enrichment of m⁶A in the brain (Extended Data Fig. 9a). Consistent with our findings, splicing of two *Sxl* target transcripts, *transformer* (*tra*) and *msl-2*, is

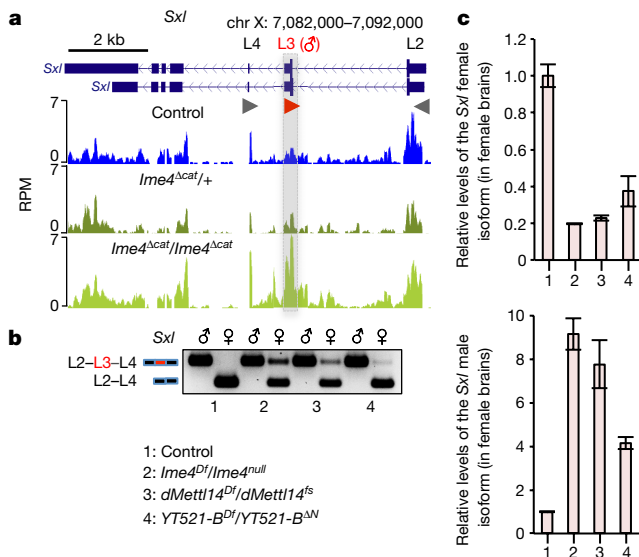


Figure 4 | YT521-B and the methyltransferase complex regulate *Sxl* splicing. **a**, RNA-seq data from control *w¹¹¹⁸*, heterozygous and homozygous *Ime4* heads. Arrowheads display the primers used for quantification in **b** and **c**. chr, chromosome; L2, late exon 2; L3, late exon 3; L4, late exon 4. **b**, Spliced isoforms for *Sxl* were monitored by semi-quantitative RT-PCR using RNA extracts from male and female heads. The genotypes used are indicated below. **c**, RT-qPCR quantification of the *Sxl* female and male isoforms from RNA extracts of female heads using primers L2 and L4 (top), as well as L2 and L3 (bottom). Error bars show mean \pm s.d. of three technical replicates from three biological experiments.

also altered (Extended Data Fig. 9b, c). These results indicate that the methyltransferase complex facilitates splicing of *Sxl* pre-mRNA, suggesting a role in sex determination and dosage compensation. To validate this hypothesis, we tested whether *Ime4* genetically interacts with *Sxl*. Transheterozygous *Ime4* females were crossed with males carrying a deficiency in the *Sxl* locus and the survival rate of the progeny was quantified for both sexes. Females lacking one wild-type copy of both *Ime4* and *Sxl* had severely reduced survival, while males were unaffected (Extended Data Fig. 9d). This effect probably arises from impairment of the dosage compensation pathway. Thus, these findings indicate that *Ime4* interacts with *Sxl* to control female survival.

Loss of YT521-B resembles m⁶A phenotypes

Given that YT521-B specifically recognizes m⁶A and influences most m⁶A-dependent splicing events in S2R+ cells, we investigated whether its deletion *in vivo* mimics the knockout of members of the methyltransferase complex. A deletion in the *YT521-B* locus that disrupts expression of both *YT521-B* isoforms was generated (Extended Data Fig. 7c). Similar to *Ime4* and *dMettl14* mutants, *YT521-B* mutant flies survive until adulthood but exhibit flight defects and poor locomotion (Fig. 3b, c and Extended Data Fig. 6e). Comparison of the transcriptome of *Ime4*-knockout with *YT521-B*-knockout female flies identified 397 splicing events regulated by *Ime4* and, among those, 243 (61% of *Ime4*-affected events) are also regulated by *YT521-B* (Extended Data Fig. 9e, f), indicating a similar overlap as from S2R+ cells. While alternative 5' splice site usage is not specifically enriched *in vivo*, intron retention is still overrepresented (Extended Data Fig. 9g). Notably, loss of YT521-B also leads to the male-specific splicing of *Sxl*, *tra* and *msl-2* and to the decrease of the female-specific *Sxl* isoform in females (Fig. 4b, c and Extended Data Fig. 9b, c). Collectively, these experiments strongly suggest that the m⁶A methyltransferase complex regulates adult locomotion and sex determination primarily via YT521-B binding to m⁶A.

Nito is a novel component of the m⁶A complex

To investigate the mechanisms of YT521-B-mediated splicing control, we searched for specific interacting partners using stable isotope labelling with amino acids in cell culture (SILAC)-based quantitative proteomics upon immunoprecipitation of a Myc-tagged YT521-B protein from S2R+ cells. We identified 73 factors that show more than twofold enrichment in the YT521-B-Myc sample (Extended Data Fig. 10a and Supplementary Table 8). Almost half ($n = 30$) are predicted mRNA-binding proteins. To investigate whether some of these mRNA-binding proteins regulate m⁶A-dependent splicing, we depleted them in S2R+ cells and assessed the effects on *fl(2)d* splicing. Notably, three proteins, Hrb27C, Qkr58E-1 and Nito, were found to similarly control *fl(2)d* splicing (Extended Data Fig. 10b). Expanding this analysis to six additional m⁶A-regulated splicing events reveals that Hrb27C and Qkr58E-1 regulate only a subset, while loss of Nito consistently leads to similar splicing defects as observed upon depletion of YT521-B and members of the methyltransferase complex (Fig. 5a and Extended Data Fig. 10c–f). To get further insights into the interplay between YT521-B and the three mRNA-binding proteins, we performed co-immunoprecipitation experiments. While Qkr58E-1 interacts with YT521-B in an RNA-independent fashion, interactions with Hrb27C and Nito could not be reproduced (Extended Data Fig. 10g, h and data not shown). However, we found that Nito interacts with both Fl(2)d and Ime4 independently of the presence of RNA (Fig. 5b). These findings indicate that Nito might be a component of the methyltransferase complex. Accordingly, *nito* mRNA expression correlates with m⁶A levels during development (Extended Data Fig. 10i) and Nito knockdown leads to a severe m⁶A decrease (Fig. 5c). This decrease is not an indirect consequence of reduced levels of methyltransferase complex components upon Nito knockdown (Extended Data Fig. 10j). Finally, YT521-B binding to mRNA depends on the presence of Nito (Fig. 5d). Collectively, these results demonstrate that Nito is a bona fide member of the m⁶A methyltransferase complex.

Discussion

Components of the methyltransferase complex were shown to be essential during early development of various organisms. In contrast to these studies, our analysis argues against a vital role for Ime4 in *Drosophila* as both our deletion alleles give rise to homozygous adults without prominent lethality during development. This cannot be explained by compensation via dMettl14, as its knockout produces similar effects as the Ime4 knockout. Furthermore, depleting both genes only slightly intensifies the locomotion phenotype without affecting fly survival, supporting the idea that Ime4 and dMettl14 act together to regulate the same target genes. Accordingly, loss of either component *in vivo* dramatically affects stability of the other (Extended Data Fig. 1i).

Loss of function of either of the methyltransferases produces severe behavioural defects. All of them can be rescued by specific expression of *Ime4* cDNA in the nervous system of *Ime4* mutants, indicating neuronal functions. This is consistent with the substantial enrichment of m⁶A and its writer proteins in the embryonic neuroectoderm, as well as with the affected genes upon depletion in S2R+ cells. Our analyses further reveal notable changes in the architecture of NMJs, potentially explaining the locomotion phenotype. In the mouse, m⁶A is enriched in the adult brain⁵, whereas in zebrafish, METTL3 and WTAP show high expression in the brain region of the developing embryo¹⁹. Furthermore, a crucial role for the mouse m⁶A demethylase FTO in the regulation of the dopaminergic pathway was clearly demonstrated²². Thus, together with previous studies, our work reveals that m⁶A RNA methylation is a conserved mechanism of neuronal mRNA regulation contributing to brain function.

We find that Ime4 and dMettl14 also control the splicing of the *Sxl* transcript, encoding for the master regulator of sex determination in *Drosophila*. This is in agreement with the previously demonstrated roles of Fl(2)d and Vir in this process^{38,39}. However, in contrast to these

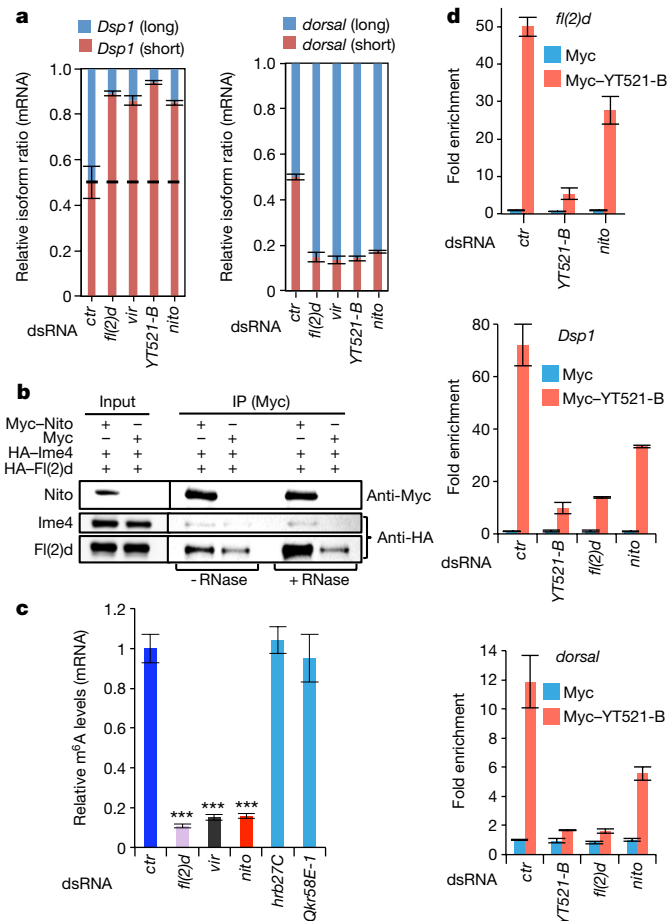


Figure 5 | Nito is a novel member of the methyltransferase complex.

a, RT-qPCR quantification of splicing isoforms of *Dsp1* and *dorsal* transcripts upon knockdown of indicated proteins. Bars charts show mean \pm s.d. of three technical replicates from two biological experiments. *ctr*, control. **b**, Lysates from S2R+ cells expressing the indicated proteins were immunoprecipitated (IP) using Myc beads. **c**, LC-MS/MS quantification of m⁶A levels in mRNA depleted for the indicated proteins. Bar chart represents the mean \pm s.d. of three technical measurements from two biological replicates. ****P* < 0.0001 (one-way ANOVA, Tukey's post-hoc analysis). **d**, RT-qPCR quantification of fold enrichment over input of m⁶A-regulated transcripts in RNA immunoprecipitation with Myc-YT521-B, after depletion of indicated proteins. Error bars represent the mean \pm s.d. of three technical measurements from two biological replicates.

mutants, mutants for *Ime4*, *dMettl14* and *YT521-B* are mostly viable, ruling out an essential role in sex determination and dosage compensation. Only when one copy of *Sxl* is removed, *Ime4* mutant females start to die. Notably, m⁶A effect on *Sxl* appears more important in the brain compared to the rest of the organism, possibly allowing fly survival in the absence of this modification.

Our targeted screen identifies Nito as a bona fide methyltransferase complex subunit. The vertebrate homologue of Nito, RBM15, was recently shown to affect *XIST* gene silencing via recruitment of the methyltransferase complex to *XIST* RNA, indicating that its role in m⁶A function and dosage compensation is conserved⁴⁰. In summary, this study provides a comprehensive *in vivo* characterization of m⁶A biogenesis and function in *Drosophila*, demonstrating the crucial importance of the methyltransferase complex in controlling neuronal functions and fine-tuning sex determination via its nuclear reader YT521-B.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 March; accepted 21 October 2016.

Published online 30 November 2016.

- Motorin, Y. & Helm, M. RNA nucleotide methylation. *Wiley Interdiscip. Rev. RNA* **2**, 611–631 (2011).
- Fu, Y., Dominissini, D., Rechavi, G. & He, C. Gene expression regulation mediated through reversible m⁶A RNA methylation. *Nature Rev. Genet.* **15**, 293–306 (2014).
- Meyer, K. D. & Jaffrey, S. R. The dynamic epitranscriptome: N⁶-methyladenosine and gene expression control. *Nature Rev. Mol. Cell Biol.* **15**, 313–326 (2014).
- Dominissini, D. *et al.* Topology of the human and mouse m⁶A RNA methylomes revealed by m⁶A-seq. *Nature* **485**, 201–206 (2012).
- Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
- Schwartz, S. *et al.* High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* **155**, 1409–1421 (2013).
- Li, Y. *et al.* Transcriptome-wide N⁶-methyladenosine profiling of rice callus and leaf reveals the presence of tissue-specific competitors involved in selective mRNA modification. *RNA Biol.* **11**, 1180–1188 (2014).
- Lin, S. & Gregory, R. I. Methyltransferases modulate RNA stability in embryonic stem cells. *Nature Cell Biol.* **16**, 129–131 (2014).
- Schwartz, S. *et al.* Perturbation of m⁶A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Reports* **8**, 284–296 (2014).
- Wang, X. *et al.* N⁶-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117–120 (2014).
- Wang, Y. *et al.* N⁶-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nature Cell Biol.* **16**, 191–198 (2014).
- Alarcón, C. R. *et al.* HNRNP2B1 is a mediator of m⁶A-dependent nuclear RNA processing events. *Cell* **162**, 1299–1308 (2015).
- Liu, N. *et al.* N⁶-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560–564 (2015).
- Meyer, K. D. *et al.* 5' UTR m⁶A promotes Cap-independent translation. *Cell* **163**, 999–1010 (2015).
- Wang, X. *et al.* N⁶-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161**, 1388–1399 (2015).
- Zhou, J. *et al.* Dynamic m⁶A mRNA methylation directs translational control of heat shock response. *Nature* **526**, 591–594 (2015).
- Wang, X. & He, C. Reading RNA methylation codes through methyl-specific binding proteins. *RNA Biol.* **11**, 669–672 (2014).
- Liu, J. *et al.* A METTL3-METTL14 complex mediates mammalian nuclear RNA N⁶-adenosine methylation. *Nature Chem. Biol.* **10**, 93–95 (2014).
- Ping, X. L. *et al.* Mammalian WTAP is a regulatory subunit of the RNA N⁶-methyladenosine methyltransferase. *Cell Res.* **24**, 177–189 (2014).
- Jia, G. *et al.* N⁶-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature Chem. Biol.* **7**, 885–887 (2011).
- Fu, Y. *et al.* FTO-mediated formation of N⁶-hydroxymethyladenosine and N⁶-formyladenosine in mammalian RNA. *Nature Commun.* **4**, 1798 (2013).
- Hess, M. E. *et al.* The fat mass and obesity associated gene (*Fto*) regulates activity of the dopaminergic midbrain circuitry. *Nature Neurosci.* **16**, 1042–1048 (2013).
- Zheng, G. *et al.* ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **49**, 18–29 (2013).
- Batista, P. J. *et al.* m⁶A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15**, 707–719 (2014).
- Geula, S. *et al.* m⁶A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science* **347**, 1002–1006 (2015).
- Hongay, C. F. & Orr-Weaver, T. L. *Drosophila* Inducer of Meiosis 4 (*IME4*) is required for Notch signaling during oogenesis. *Proc. Natl Acad. Sci. USA* **108**, 14855–14860 (2011).
- Zhong, S. *et al.* MTA is an *Arabidopsis* messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* **20**, 1278–1288 (2008).
- Bodi, Z. *et al.* Adenosine methylation in *Arabidopsis* mRNA is associated with the 3' end and reduced levels cause developmental defects. *Front. Plant Sci.* **3**, 48 (2012).
- Shah, J. C. & Clancy, M. J. *IME4*, a gene that mediates MAT and nutritional control of meiosis in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **12**, 1078–1086 (1992).
- Clancy, M. J., Shambaugh, M. E., Timpte, C. S. & Bokar, J. A. Induction of sporulation in *Saccharomyces cerevisiae* leads to the formation of N⁶-methyladenosine in mRNA: a potential mechanism for the activity of the *IME4* gene. *Nucleic Acids Res.* **30**, 4509–4518 (2002).
- Agarwala, S. D., Blitzblau, H. G., Hochwagen, A. & Fink, G. R. RNA methylation by the MIS complex regulates a cell fate decision in yeast. *PLoS Genet.* **8**, e1002732 (2012).
- Wang, P., Doxtader, K. A. & Nam, Y. Structural basis for cooperative function of Mett13 and Mett14 methyltransferases. *Mol. Cell* **63**, 306–317 (2016).
- Wang, X. *et al.* Structural basis of N⁶-adenosine methylation by the METTL3-METTL14 complex. *Nature* **534**, 575–578 (2016).
- Gabel, H. W. *et al.* Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**, 89–93 (2015).
- Götz, K. G. Visual guidance in *Drosophila*. *Basic Life Sci.* **16**, 391–407 (1980).

36. Strauss, R., Hanesch, U., Kinkelin, M., Wolf, R. & Heisenberg, M. No-bridge of *Drosophila melanogaster*: portrait of a structural brain mutant of the central complex. *J. Neurogenet.* **8**, 125–155 (1992).
37. Clough, E. & Oliver, B. Genomics of sex determination in *Drosophila*. *Brief. Funct. Genomics* **11**, 387–394 (2012).
38. Granadino, B., Campuzano, S. & Sánchez, L. The *Drosophila melanogaster* fl(2)d gene is needed for the female-specific splicing of Sex-lethal RNA. *EMBO J.* **9**, 2597–2602 (1990).
39. Hilfiker, A., Amrein, H., Dübendorfer, A., Schneiter, R. & Nöthiger, R. The gene virilizer is required for female-specific splicing controlled by Sxl, the master gene for sexual development in *Drosophila*. *Development* **121**, 4017–4026 (1995).
40. Patil, D. P. *et al.* m⁶A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank the Bloomington *Drosophila* Stock Center for fly reagents; the *Drosophila* Genomics Resource Center at Indiana University for plasmids; The Developmental Studies Hybridoma Bank and the Lehmann laboratory for antibodies; M. Soller for sharing unpublished data; members of the J.-Y.R. laboratory for helpful discussion; A. Dold and V. Morin

for experimental help; the IMB Genomics; the IMB Genomics, Proteomics and Bioinformatics Core facilities for support; and Bioinformatics Core facilities for support; and R. Ketting, N. Soshnikova, R. Strauss, J. Treisman and K. Zarnack for critical reading of the manuscript. Research in the laboratory of J.-Y.R. is supported by the Marie Curie CIG 334288 and the Deutsche Forschungsgemeinschaft (DFG) SPP1935 grant RO 4681/4-1. L.S. is funded by the Rhineland-Palatinate program Gene RED. The project was also supported by a DFG grant (HE 3397/13-1, SPP1784) to M.H.

Author Contributions T.L. and J.-Y.R. conceived the idea. T.L. designed and performed the experiments. J.A. performed the YT521-B RNA immunoprecipitation experiment and M.B. generated the YT521-B allele. K.S. and M.H. performed the LC-MS/MS quantification of m⁶A levels. L.S. and B.P. carried out the Buridan analysis. C.H.H. performed NMJ staining and analysis. M.A.A.-N. performed the phylogenetic analysis. N.K. performed the computational analysis. T.L. and J.-Y.R. wrote the manuscript with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.-Y.R. (j.roignant@imb-mainz.de).

METHODS

Drosophila stocks and genetics. *Drosophila melanogaster* *w*¹¹¹⁸, Canton-S and Oregon-R were used as wild-type controls. Other fly stocks used were *tub-GAL4*, *elav*^{C155}-GAL4, *how*^{24B}-GAL4, *Df(3R)Exel6197*, *Df(2L)BSC200/Cyo*, *Df(3L)Exel6094* (Bloomington *Drosophila* Stock Center). UAS-Ime4-HA/Cyo flies were generated by injection of UAS Ime4-HA vector at Bestgene. Mutant alleles for *Ime4*, *dMettl14*, *fl(2)d* and *YT521-B* were generated using the CRISPR-Cas9 system following the previously described procedure⁴¹. Two independent guide RNAs (gRNAs) per gene were designed using the gRNA design tool: <http://www.crisprflydesign.org/> (Supplementary Table 9). Oligonucleotides were annealed and cloned into pBfV-U6.2 vector (National Institute of Genetics, Japan). Vectors were injected into embryos of *y² cho² v¹*; *attP40(U6.2-w-ex3-2)* flies. Positive recombinant males were further crossed with *y² cho² v¹*; *attP40(nos-Cas9)/Cyo* females. Males carrying *nos-Cas9* and U6-gRNA transgenes were screened for the expected deletion and further crossed with the balancer strain *Apt^{Ex}/CyoGFP-TM6c*. *Ime4*^{Δcat} allele was obtained using gRNA sequences (GGACTCTTTCCGCGCTACAG and GGCTCACACGGACGAATCTC). A deletion of 569 bp (607–1,175 bp in the genome region chr3R:24032157..24034257, genome assembly BDGP release 6) was produced. *Ime4*^{null} allele was obtained using gRNA sequences (GGCCCTTTTAACGTTCTTGA and GGCTCACACGGACGAATCTC) and produced a deletion of 1,291 bp (1,876–3,166 bp in the genome region chr3R:24030157..24034257). *dMettl14*^Δ allele was obtained using gRNA sequences (GGTTCCTTCAGGAAGGTCG and GGACCAACATTAACAAGCCC) and produced a 2-nucleotide frame shift at position 227 of the coding sequence, leading to a premature stop codon at amino acid position 89. *YT521-B*^{ΔN} allele was obtained using gRNA sequences (GGCATTAATTGTGTGGACAC and GGCTGTCGATCCTCGGTATC) and produced a deletion of 602 bp (133–734 bp in the genome region chr3L:3370451..3374170 (reverse complemented)).

Phylogenetic analysis. The phylogenetic trees were constructed with ClustalX⁴² from multiple sequence alignments generated with MUSCLE⁴³ of the *Drosophila* sequences with homologues from representative species.

Cell line. *Drosophila* S2R+ are embryonic-derived cells obtained from *Drosophila* Genomics Resource Center (DGRC; FlyBase accession FBtc0000150). The presence of Mycoplasma contamination was not tested.

Cloning. The plasmids used for immunohistochemistry and co-immunoprecipitation assays in *Drosophila* S2R+ cells were constructed by cloning the corresponding cDNA in the pPAC vector⁴⁴ with N-terminal Myc tag and the Gateway-based vectors with N-terminal Flag-Myc tag (pPFMW) as well as C-terminal HA tag (pPWH) (obtained from *Drosophila* Genomics Resource Center at Indiana University).

Climbing test. Two-to-three-day-old flies were gender-separated and placed into measuring cylinders to assess their locomotion using the climbing assay reported previously⁴⁵. Flies were tapped to the bottom and the number of flies that climb over the 10 cm threshold in 10 s interval were counted. Ten female flies were used per experiment and six independent measurements were performed.

Staging. Staging experiment was performed using *Drosophila melanogaster* *w*¹¹¹⁸ flies that were kept in a small fly cage at 25 °C. Flies laid embryos on big apple juice plates that were exchanged every 2 h. Before each start of collection, 1 h pre-laid embryos were discarded to remove all retained eggs and embryos from the collection. All the resultant plates with embryos of 1 h or 2 h lay were further incubated at 25 °C between 0 h and 20 h, with 2 h increments, to get all embryonic stages. For the collection of larval stages, L1 larvae (~30 larvae/stage) were transferred onto a new apple juice plate and were further incubated at 25 °C till they reached a defined age (24 to 110 h, 2 h intervals). Similarly, pupal stages were obtained by the transfer of L3 larvae (~30/stage) in a fresh vial, that were kept at 25 °C and left to develop into defined stage between 144 and 192 h in 2 h increments. One-to-three-day-old adults were collected and gender separated. Heads and ovaries from 50 females were also collected. A total of three independent samples were collected for each *Drosophila* stage as well as for heads and ovaries. Samples from the staging experiment were used for RNA extraction to analyse m⁶A abundance in mRNA and expression levels of different transcripts during *Drosophila* development.

RNA isolation and measurement of RNA levels. Total RNA from S2R+ cells was isolated using Trizol reagent (Invitrogen) and DNA was removed with DNase-I treatment (NEB). mRNA was purified with Oligotex mRNA Kit (Qiagen) or by using two rounds of purification with Dynabeads Oligo (dT)25 (Invitrogen). cDNA for RT-qPCR was prepared using M-MLV Reverse Transcriptase (Promega) and transcript levels were quantified using Power SYBR Green PCR Master Mix (Invitrogen) and the oligonucleotides indicated in Supplementary Table 9.

For RNA isolation from fly heads, 20 female flies were collected in 1.5 ml Eppendorf tubes and flash frozen in liquid nitrogen. Heads were first removed from the body by spinning the flies on vortex and then collected via the 0.63 mm sieve at 4 °C. Fly heads were homogenized using a pestle and total RNA was isolated

with Trizol reagent. DNA was removed by DNase-I treatment and RNA was further purified using RNeasy Kit (Qiagen). RNA from adult flies and dissected ovaries was prepared as described earlier by skipping the head separation step.

RT-PCR. Two-to-three-day-old flies were collected and their RNA isolated as described earlier. Following cDNA synthesis PCR was performed using the oligonucleotides described in Supplementary Table 9 to analyse *Sxl*, *tra* and *msl-2* splicing.

RNA *in situ* hybridization. For *in situ* hybridization *Drosophila melanogaster* *w*¹¹¹⁸ flies were kept at 25 °C in conical flasks with apple juice agar plates and embryos were collected every 24 h. Embryos were transferred in a sieve and dechorionated for 2 min in 50% sodium hypochloride. After 5 min wash in water, embryos were permeabilized with PBST (0.1% Tween X-100 in PBS) for 5 min. Embryos were transferred in 1:1 mixture of heptane (Sigma) and 8% formaldehyde (Sigma) and fixed for 20 min with constant shaking at room temperature. After fixation the lower organic phase was removed and 1 volume of MeOH was added to the aqueous phase containing fixed embryos. Following 5 min of extensive shaking all liquid was removed and embryos were washed 3 times with 100% MeOH. At this point embryos were stored at –20 °C or used for further analysis. For *in situ* hybridization MeOH was gradually replaced with PBST with 10 min washes and with three final washes in PBST. Embryos were further washed for 10 min at room temperature with 50% HB4 solution (50% formamide, 5× SSC, 50 μg/ml heparin, 0.1% Tween, 5 mg/ml torula yeast extract) diluted in PBST. Blocking was performed with HB4 solution, first for 1 h at room temperature and next for 1 h at 65 °C. *In situ* probes were prepared with DIG RNA labelling Kit (Roche) following the manufacturer's protocol. Two microlitres of the probe were diluted in 200 μl of HB4 solution, heated up to 65 °C to denature the RNA secondary structure and added to blocked embryos for further overnight incubation at 65 °C. The next day, embryos were washed 2 times for 30 min at 65 °C with formamide solution (50% formamide, 1× SSC in PBST) and further 3 times for 20 min at room temperature with PBST. Embryos were then incubated with anti-DIG primary antibody (Roche) diluted in PBST (1:2,000) for 2 h at room temperature and later washed 5 times for 30 min with PBST. In order to develop the staining, embryos were rinsed with AP buffer (100 mM Tris pH 9.5, 50 mM MgCl₂, 100 mM NaCl, 0.1% Tween) and incubated with NBT/BCIP solution in AP buffer (1:100 dilution) until the intense staining was observed. Reaction was stopped with several 15 min PBST washes. Prior to mounting, embryos were incubated in 20% glycerol and later visualized on Leica M205-FA stereomicroscope.

YT521-B RNA immunoprecipitation. S2R+ cells were depleted for the indicated proteins with two treatments of double-stranded RNA (dsRNA). Four days after treatment Myc-tagged YT521-B was transfected along with the control Myc construct. Seventy-two hours after transfection, cells were fixed with 1% formaldehyde at room temperature for 10 min and harvested as described previously⁴⁶. Extracted nuclei were subjected to 13 cycles of sonication on a bioruptor (Diagnode), with 30 s “ON”/“OFF” at high settings. Nuclear extracts were incubated overnight with 4 μg of anti-Myc 9E10 antibody (Enzo Life Sciences). Immunoprecipitation was performed as described previously⁴⁶ except that samples were DNase-treated (NEB) instead of RNase-treated and subjected to proteinase K treatment for reversal of crosslinks, 1 h at 65 °C. RNase inhibitors (Murine RNase Inhibitor, NEB) were used in all steps of the protocol at a concentration of 40 U/ml.

Generation of antibodies. Antibodies against Ime4 and dMettl14 were generated at Eurogentec. For anti-Ime4 sera guinea pig was immunized with a 14 amino-acid-long peptide (163–177 amino acids (AA)); for anti-dMettl14 sera rabbit was immunized with a 14 amino acid-long peptide (240–254 AA). Both serums were affinity-purified using peptide antigens crosslinked to sepharose columns.

Immunostaining. For ovary immunostaining, ovaries from 3–5-day-old females were dissected in ice-cold PBS and fixed in 5% formaldehyde for 20 min at room temperature. After a 10 min wash in PBT1% (1% Triton X-100 in PBS), ovaries were further incubated in PBT1% for 1 h at room temperature. Ovaries were then blocked with PBTB (0.2% Triton, 1% BSA in PBS) for 1 h at room temperature and later incubated with the primary antibodies in PBTB overnight at 4 °C: rabbit anti-Vasa, 1:250 (gift from Lehmann laboratory), mouse anti-ORB 1:30 (#6H4 DSHB). The following day, ovaries were washed 2 times for 30 min in PBTB and blocked with PBTB containing 5% donkey serum (Abcam) for 1 h at room temperature. Secondary antibody was added later in PBTB with donkey serum and ovaries were incubated for 2 h at room temperature. Five washing steps of 30 min were performed with 0.2% Triton in PBT and ovaries were mounted onto slides in Vectashield (Vector Labs).

For NMJ staining, third instar larvae were dissected in calcium free HL-3 saline and fixed in 4% paraformaldehyde in PBT (PBS + 0.05% Triton X-100). Larvae were then washed briefly in 0.05% PBT for 30 min and incubated overnight at 4 °C with the following primary antibodies: rabbit anti-synaptotagmin, 1:2,000

(ref. 47); mouse anti-DLG, 1:100 (#4F3, DSHB); TRITC-conjugated anti-HRP, 1:200 (Jackson ImmunoResearch). Secondary antibodies conjugated to Alexa-488 (goat anti-rabbit, Jackson ImmunoResearch) and Alexa-647 (goat anti-mouse, Jackson ImmunoResearch) were used at a concentration of 1:200 and incubated at room temperature for 2 h. Larvae were finally mounted in Vectashield.

For staining of *Drosophila* S2R+ cells, cells were transferred to the poly-lysine pre-treated 8-well chambers (Ibidi) at the density of 2×10^5 cells/well. After 30 min, cells were washed with $1 \times$ DPBS (Gibco), fixed with 4% formaldehyde for 10 min and permeabilized with PBST (0.2% Triton X-100 in PBS) for 15 min. Cells were incubated with mouse anti-Myc (1:2000; #9E10, Enzo) in PBST supplemented with 10% of donkey serum at 4°C, overnight. Cells were washed $3 \times$ for 15 min in PBST and then incubated with secondary antibody and $1 \times$ DAPI solution in PBST supplemented with 10% of donkey serum for 2 h at 4°C. After three 15 min washes in PBST, cells were imaged with Leica SP5 confocal microscope using $\times 63$ oil immersion objective.

NMJ analysis. Images from muscles 6–7 (segment A3) were acquired with a Leica Confocal Microscope SP5. Serial optical sections at 512×512 or $1,024 \times 1,024$ pixels were obtained at $0.38 \mu\text{m}$ with the $\times 63$ objective. Different genotypes were processed simultaneously and imaged using identical confocal acquisition parameters for comparison. Bouton number was quantified in larval abdominal segment A3, muscles 6 and 7, of wandering third instar larvae. ImageJ software (version 1.49) was used to measure the area of the synaptotagmin-positive area.

Cell culture, RNA interference and transfection. *Drosophila* S2R+ cells were grown in Schneider's medium (Gibco) supplemented with 10% FBS (Sigma) and 1% penicillin–streptomycin (Sigma). For RNA interference (RNAi) experiments, PCR templates for the dsRNA were prepared using T7 megascript Kit (NEB). dsRNA against bacterial β -galactosidase gene (*lacZ*) was used as a control for all RNA interference (RNAi) experiments. S2R+ cells were seeded at the density of 10^6 cells/ml in serum-free medium and $7.5 \mu\text{g}$ of dsRNA was added to 10^6 cells. After 6 h of cell starvation, serum supplemented medium was added to the cells. dsRNA treatment was repeated after 48 and 96 h and cells were collected 24 h after the last treatment. Effectene (Qiagen) was used to transfect vector constructs in all overexpression experiments following the manufacturer's protocol.

Co-immunoprecipitation assay and western blot analysis. For the co-immunoprecipitation assay, different combinations of vectors with indicated tags were co-transfected in S2R+ cells seeded in a 10 cm cell culture dish as described earlier. Forty-eight hours after transfection cells were collected, washed with DPBS and pelleted by 10 min centrifugation at 400g. The cell pellet was lysed in 1 ml of lysis buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.05% NP-40) supplemented with protease inhibitors (1 $\mu\text{g}/\text{ml}$ leupeptin, 1 $\mu\text{g}/\text{ml}$ pepstatin, 1 $\mu\text{g}/\text{ml}$ aprotinin and 1 mM PMSF) and rotated head-over-tail for 30 min at 4°C. Nuclei were collected by 10 min centrifugation at 1,000g at 4°C re-suspended in 300 μl of lysis buffer and sonicated with 5 cycles of 30 s ON, 30 s OFF low power setting. Cytoplasmic and nuclear fractions were joined and centrifuged at 18,000g for 10 min at 4°C to remove the remaining cell debris. Protein concentrations were determined using Bradford reagent (BioRad). For immunoprecipitation, 2 mg of proteins were incubated with 7 μl of anti-Myc antibody coupled to magnetic beads (Cell Signaling) in lysis buffer and rotated head-over-tail overnight at 4°C. The beads were washed 3 times for 15 min with lysis buffer and immunoprecipitated proteins were eluted by incubation in $1 \times$ NuPAGE LDS buffer (ThermoFischer) at 70°C for 10 min. Eluted immunoprecipitated proteins were removed from the beads and DTT was added to 10% final volume. Immunoprecipitated proteins and input samples were analysed by western blot after incubation at 70°C for additional 10 min.

For western blot analysis, proteins were separated on 7% SDS–PAGE gel and transferred on Nitrocellulose membrane (BioRad). After blocking with 5% milk in PBST (0.05% Tween in PBS) for 1 h at room temperature, the membrane was incubated with primary antibody in blocking solution overnight at 4°C. Primary antibodies used were: mouse anti-Myc 1:2,000 (#9E10, Enzo); mouse anti-HA 1:1,000 (#16B12, COVANCE); mouse anti-Tubulin 1:2,000 (#903401, Biolegend); guinea pig anti-Ime4 1:500 and rabbit anti-dMettl14 1:250. The membrane was washed 3 times in PBST for 15 min and incubated 1 h at room temperature with secondary antibody in blocking solution. Protein bands were detected using SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific).

In vitro pull-down assay. S2R+ cells were transfected with either Myc–YT521-B or Myc–GFP constructs. Forty-eight hours after transfection cells were collected, washed with PBS and pelleted by centrifugation at 400g for 10 min. The cell pellet was lysed and processed in 1 ml of pull-down lysis buffer (10 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2 mM EDTA, 0.5% Triton-X100, 0.5 mM DTT). For individual pull-down, 1.5 mg of protein were incubated with either 3 μg of biotinylated RNA probe of *bPRL* containing m⁶A or not, or without probe, as a control in 0.5 ml of pull-down buffer supplemented with protease inhibitor mix and 10 U of Murine RNase Inhibitor (NEB) and incubated for 2 h at 4°C. Five microlitres

of Streptavidin beads (M-280, Invitrogen) were added and pull-down samples were incubated for an additional 1 h at 4°C. After 3 washes of 15 min with pull-down buffer, beads were re-suspended in 400 μl of pull-down buffer. One-hundred microlitres was used for RNA isolation and dot blot analysis of recovered RNA probes with anti-Strep-HRP. The remaining 300 μl of the beads was collected on the magnetic rack and immunoprecipitated proteins were eluted by incubation in $1 \times$ SDS buffer (ThermoFischer) at 95°C for 10 min. Immunoprecipitated proteins as well as input samples were analysed by western blot.

Dot blot assays. Serial dilutions of biotinylated RNA probe of *bPRL* containing m⁶A or A were spotted and crosslinked on nitrocellulose membrane (Biorad) with ultraviolet 245 light ($3 \times 150 \text{ mJ}/\text{cm}^2$). RNA loading was validated with methylene blue staining. Membranes were blocked with 5% milk in PBST for 1 h at room temperature and washed in PBST before incubation with the proteins of interest. S2R+ cells were transfected with either Myc–YT521-B or Myc–GFP constructs. Forty-eight hours after transfection cells were collected, washed with PBS and pelleted by centrifugation at 400g for 10 min. The cell pellet was lysed in 1 ml of lysis buffer (10 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.5% NP-40). Three milligrams of the protein lysate were mixed with 2% BSA in lysis buffer and incubated with the membrane overnight at 4°C. For control dot-blot rabbit anti-m6A antibody (Synaptic Systems) was used. The next day membranes were washed 3 times in lysis buffer. Membranes with bound proteins were further crosslinked with ultraviolet 245 light ($3 \times 150 \text{ mJ}/\text{cm}^2$) and analysed using anti-Myc antibody.

SILAC experiment and LC–MS/MS analysis. For SILAC experiments, S2R+ cells were grown in Schneider medium (Dundee Cell) supplemented with either heavy (Arg8, Lys8) or light amino acids (Arg0, Lys0) (Sigma). For the forward experiment, Myc–YT521-B was transfected in heavy-labelled cells and Myc-alone in light-labelled cells. The reverse experiment was performed vice versa. The co-immunoprecipitation experiment was done as described earlier. Before elution, beads of the heavy and light lysates were combined in 1:1 ratio and eluted with $1 \times$ NuPAGE LDS buffer that was subject to MS analysis as described previously⁴⁸. Raw files were processed with MaxQuant (version 1.5.2.8) and searched against the Uniprot database of annotated *Drosophila* proteins (*Drosophila melanogaster*: 41,850 entries, downloaded 8 January 2015).

LC–MS/MS analysis of m⁶A levels. mRNA samples from S2R+ cells depleted for the indicated proteins or from *Drosophila* staging experiments were prepared following the aforementioned procedure. Three-hundred nanograms of purified mRNA was further digested using 0.3 U Nuclease P1 from *Penicillium citrinum* (Sigma-Aldrich, Steinheim, Germany) and 0.1 U Snake venom phosphodiesterase from *Crotalus adamanteus* (Worthington, Lakewood, USA). RNA and enzymes were incubated in 25 mM ammonium acetate, pH 5, supplemented with 20 μM zinc chloride for 2 h at 37°C. Remaining phosphates were removed by 1 U FastAP (Thermo Scientific, St Leon-Roth, Germany) in a 1 h incubation at 37°C in the manufacturer supplied buffer. The resulting nucleoside mix was then spiked with ¹³C stable isotope labelled nucleoside mix from *Escherichia coli* RNA as an internal standard (SIL-IS) to a final concentration of 6 ng/ μl for the sample RNA and 10 ng/ μl for the SIL-IS. For analysis, 10 μl of the before mentioned mixture were injected into the LC–MS/MS machine. Generation of technical triplicates was obligatory. All mRNA samples were analysed in biological triplicates, except for the *ctr*, *mito*, *vir*, *hrb27C* and *qkr58E-1* knockdown samples, which were measured as biological duplicates.

LC separation was performed on an Agilent 1200 series instrument, using 5 mM ammonium acetate buffer as solvent A and acetonitrile as buffer B. Each run started with 100% buffer A, which was decreased to 92% within 10 min. Solvent A was further reduced to 60% within another 10 min. Until minute 23 of the run, solvent A was increased to 100% again and kept at 100% for 7 min to re-equilibrate the column (Synergi Fusion, 4 μM particle size, 80 Å pore size, $250 \times 2.0 \text{ mm}$, Phenomenex, Aschaffenburg, Germany). The ultraviolet signal at 254 nm was recorded via a DAD detector to monitor the main nucleosides.

MS/MS was then conducted on the coupled Agilent 6460 Triple Quadrupole (QQQ) mass spectrometer equipped with an Agilent JetStream ESI source which was set to the following parameters: gas temperature, 350°C; gas flow, 8 l/min; nebulizer pressure, 50 psi; sheath gas temperature, 350°C; sheath gas flow, 12 l/min; and capillary voltage, 3,000 V. To analyse the mass transitions of the unlabelled m⁶A and all ¹³C m⁶A simultaneously, we used the dynamic multiple reaction monitoring mode. Mass transitions, retention times and QQQ parameters are listed in Supplementary Table 10.

The quantification was conducted as described previously⁴⁹. Briefly, the amount of adenosine was evaluated by the external linear calibration of the area under the curve (AUC) of the ultraviolet signal. The amount of modification was calculated by linear calibration of the SIL-IS in relation to m⁶A nucleoside. The R^2 of both calibrations was at least 0.998 (see Extended Data Fig. 1a, b). Knowing both amounts, the percentage of m⁶A/A could be determined.

MeRIP. MeRIP was performed using the previously described protocol⁵⁰ with the following modifications. Eight micrograms of purified mRNA from *Drosophila* S2R+ cells was incubated with 5 µg of anti-m⁶A antibody (Synaptic Systems) in MeRIP buffer (150 mM NaCl, 10 mM Tris-HCl pH 7.4, 0.1% NP-40) supplemented with 5 U/ml of Murine RNase inhibitor (NEB) for 2 h at 4 °C. In control MeRIP experiment, no antibody was used in the reaction mixture. Five microlitres of A+G magnetic beads were added to all MeRIP samples for 1 h at 4 °C. On bead digestion with RNase T1 (Thermo Fisher) at final concentration 0.1 U/ml was performed for 15 min at room temperature. Beads with captured RNA fragments were then immediately washed 3 times with 500 µl of ice-cold MeRIP buffer and further eluted with 100 µl of elution buffer (0.02 M DTT, 150 mM NaCl, 50 mM Tris-HCl pH 7.4, 1 mM EDTA, 0.1% SDS, 5 U/ml Proteinase K) at 42 °C for 5 min. Elution step was repeated 4 times and 500 µl of acidic phenol/chloroform pH 4.5 (Ambion) was added to 400 µl of the combined eluate per sample to extract captured RNA fragments. Samples were mixed and transferred to Phase Lock Gel Heavy tubes (5Prime) and centrifuged for 5 min at 12,000g. Aqueous phase was precipitated overnight, −80 °C. On the following day, samples were centrifuged, washed twice with 80% EtOH and re-suspended in 10 µl of RNase-free H₂O (Ambion). Recovered RNA was analysed on RNA Pico Chip (Agilent) and concentrations were determined with RNA HS Qubit reagents. Since no RNA was recovered in the MeRIP control samples, libraries were prepared with 30 ng of duplicate MeRIPs and duplicate input mRNA samples. MeRIP-qPCR was performed on the fraction of eluted immunoprecipitated RNA and an equal amount of input mRNA. cDNA for RT-qPCR was prepared using M-MLV Reverse Transcriptase (Promega) and transcript levels were quantified using Power SYBR Green PCR Master Mix (Invitrogen) using oligonucleotides indicated in Supplementary Table 9.

Lifespan assay. For lifespan assay, 2–3-day-old flies were gender-separated and kept at 25 °C in flasks with apple juice medium (<20 flies/tube). Number of flies tested: females (37, *Ime4*^{Δcat}/*Ime4*^{Δcat}; 57, *Tubulin-GAL4/UAS-Ime4*); males (33, *Ime4*^{Δcat}/*Ime4*^{Δcat}; 41, *Tubulin-GAL4/UAS-Ime4*). To monitor their survival rate over time, flies were counted and transferred into a new tube every 2 days.

Buridan paradigm. Behavioural tests were performed on 2–5-day-old females with *Canton-S* as wild-type control. Wings were cut under cold anaesthesia to one-third of their length on the evening before the experiment. Walking and orientation behaviour was analysed using Buridan's paradigm as described³⁶. Dark vertical stripes of 12° horizontal viewing angle were shown on opposite sides of an 85-mm diameter platform surrounded by water. The following parameters were extracted by a video-tracking system (5 Hz sampling rate): total fraction of time spent walking (activity), mean walking speed taken from all transitions of a fly between the visual objects, and number of transitions between the two stripes. The visual orientation capacity (mean angular deviation) of the flies was assessed by comparing all 0.2-s path increments per fly (4,500 values in 15 min) to the respective direct path towards the angular-wise closer of the two dark stripes. All statistical groups were tested for normal distribution with the Shapiro–Wilk test. Multiple comparisons were performed using the Kruskal–Wallis ANOVA or one-way ANOVA with a post-hoc Bonferroni correction. *n* = 15 for all genotypes. The sample size was chosen based on a previous study⁵¹ and its power was validated with result analysis. Blinding was applied during the experiment.

RNA-seq. For samples from S2R+ cells and for full fly RNA samples, Illumina TruSeq Sequencing Kit (Illumina) was used. For *Drosophila* head samples, NEBNext Ultra Directional RNA Kit (NEB) was used. Libraries were prepared following the manufacturer's protocol and sequenced on Illumina HiSeq 2500. The read-length was 71 bp paired end. For MeRIP, NEBNext Ultra Directional Kit was used omitting the RNA fragmentation step for recovered MeRIP samples and following the manufacturer's protocol for input samples. Libraries were sequenced on an Illumina MiSeq as 68 bp single read in one pool on two flow cells.

Computational analysis. The RNA-seq data was mapped against the *Drosophila* genome assembly BDGP6 (Ensembl release 79) using STAR⁵² (version 2.4.0). After mapping, the bam files were filtered for secondary alignments using samtools (version 1.2). Reads on genes were counted using htseq-count (version 0.6.1p1). After read counting, differential expression analysis was done between conditions using DESeq2 (version 1.6.3) and filtered for a false discovery rate (FDR) < 5%. Differential splicing analysis was performed using rMATS (3.0.9) and filtered for FDR < 10%. The data from fly heads were treated as above but cleaned for

mitochondrial and rRNA reads after mapping before further processing. The sample *Ime4hom_3* was excluded as an outlier from differential expression analysis.

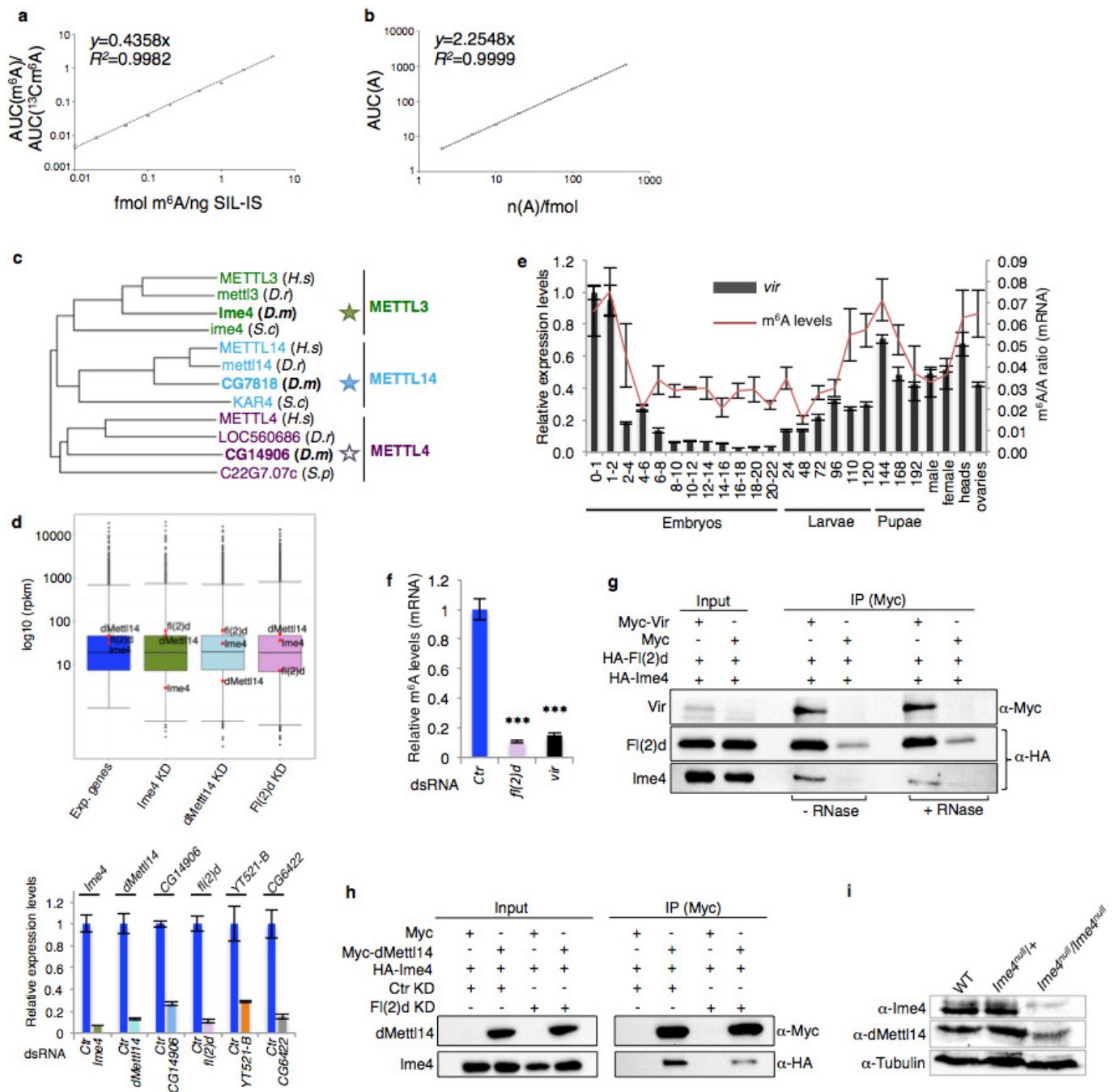
The MeRIP-seq data were processed following the same protocol as the RNA samples for mapping and filtering of the mapped reads. After mapping, peaks were called using MACS (version 1.4.1)⁵³. The genome size used for the MACS was adjusted to reflect the mappable transcriptome size based on Ensembl-annotated genes (Ensembl release 79). After peak calling, peaks were split into subpeaks using PeakSplitter (version 1.0, <http://www.ebi.ac.uk/research/bertone/software>). Consensus peaks were obtained by intersecting subpeaks of both replicates (using BEDTools, version 2.25.0). For each consensus peak, the coverage was calculated as counts per million (CPM) for each of the samples and averaged for input and MeRIP samples. Fold changes for MeRIP over input were calculated based on these. Peaks were filtered for a minimal fold change of 1.3 and a minimal coverage of 3 CPM in at least one of the samples. Peaks were annotated using the ChIPseeker and the GenomicFeatures package (based on R/Bioconductor)⁵⁴.

Statistics. In the Buridan paradigm, normality was tested for every dataset; different tests were used depending on the outcome. For not normally distributed data, Kruskal–Wallis test and Wilcoxon test were used. For normally distributed data, Bartlett test was applied to check for homogeneity of variances. ANOVA and *t*-test were used. Bonferroni corrections were applied. For climbing assays, normality was tested for every dataset. Homogeneity of variances were analysed with Levene's test. One-way ANOVA test with Tukey's post-hoc test was performed for multiple comparisons and Student's *t*-test when two data sets were compared. For m⁶A level measurement, normality was tested for every dataset. Homogeneity of variances were analysed with Levene's test. One-way ANOVA test with Tukey's post-hoc test was performed for multiple comparisons.

Randomization. Randomization was used for selection of female flies of chosen genotype for climbing tests, Buridan paradigm and RNA sequencing experiments. Randomized complete block design was applied to ensure the equal number of flies per test group. Complete randomization was applied for selection of larvae or flies of the chosen genotype for lifespan assay and NMJ staining experiment.

Data availability statement. The data that support the findings of this study have been deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE79297. All other relevant data are available from the corresponding author.

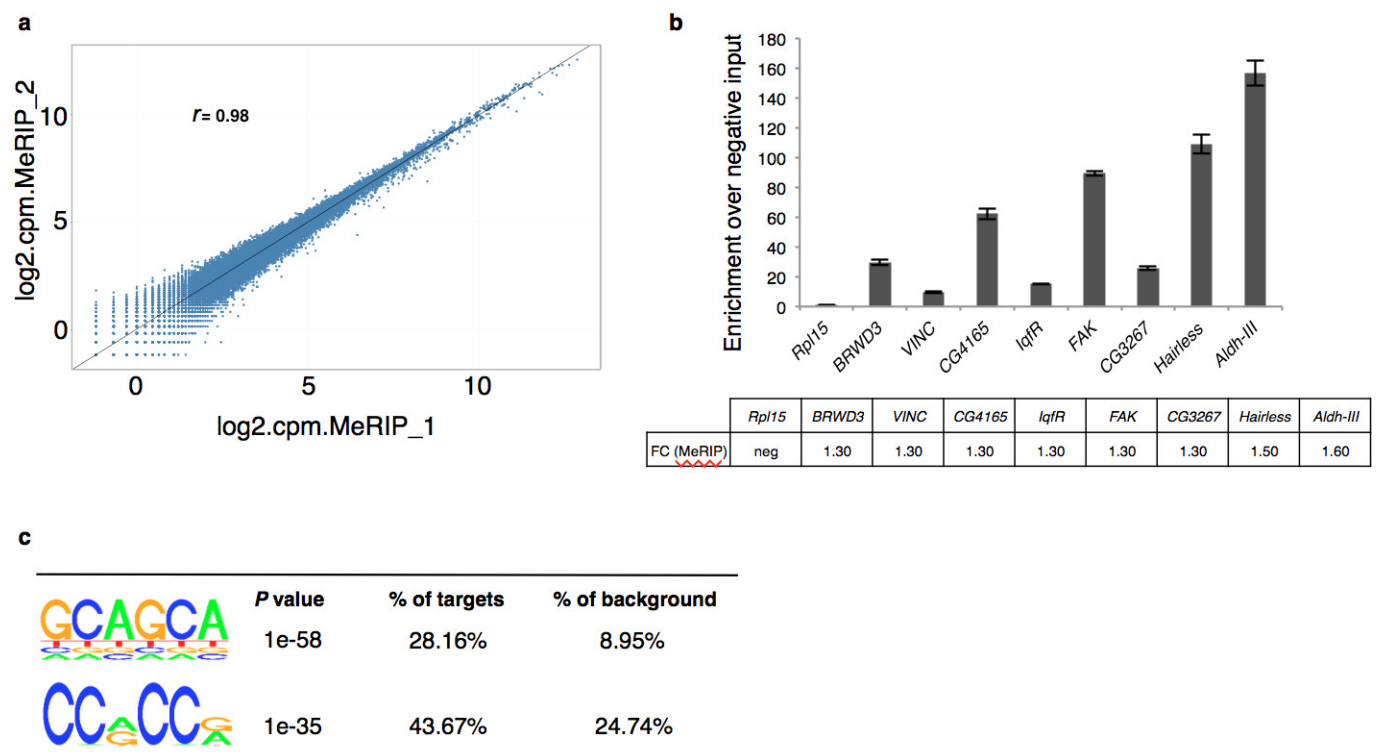
- Kondo, S. & Ueda, R. Highly improved gene targeting by germline-specific Cas9 expression in *Drosophila*. *Genetics* **195**, 715–721 (2013).
- Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
- Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
- Astigarraga, S., Hofmeyer, K., Farajian, R. & Treisman, J. E. Three *Drosophila* liprins interact to control synapse formation. *J. Neurosci.* **30**, 15358–15368 (2010).
- Bahadorani, S. & Hilliker, A. J. Cocoa confers life span extension in *Drosophila melanogaster*. *Nutr. Res.* **28**, 377–382 (2008).
- Oktaba, K. *et al.* Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in *Drosophila*. *Dev. Cell* **15**, 877–889 (2008).
- Littleton, J. T., Bellen, H. J. & Perin, M. S. Expression of synaptotagmin in *Drosophila* reveals transport and localization of synaptic vesicles to the synapse. *Development* **118**, 1077–1088 (1993).
- Bluhm, A., Casas-Vila, N., Scheibe, M. & Butter, F. Reader interactome of epigenetic histone marks in birds. *Proteomics* **16**, 427–436 (2016).
- Kellner, S. *et al.* Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Res.* **42**, e142 (2014).
- Deng, X. *et al.* Widespread occurrence of N⁶-methyladenosine in bacterial mRNA. *Nucleic Acids Res.* **43**, 6557–6567 (2015).
- Poeck, B., Triphan, T., Neuser, K. & Strauss, R. Locomotor control by the central complex in *Drosophila*—An analysis of the *tay bridge* mutant. *Dev. Neurobiol.* **68**, 1046–1058 (2008). 10.1002/dneu.20643
- Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
- Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).



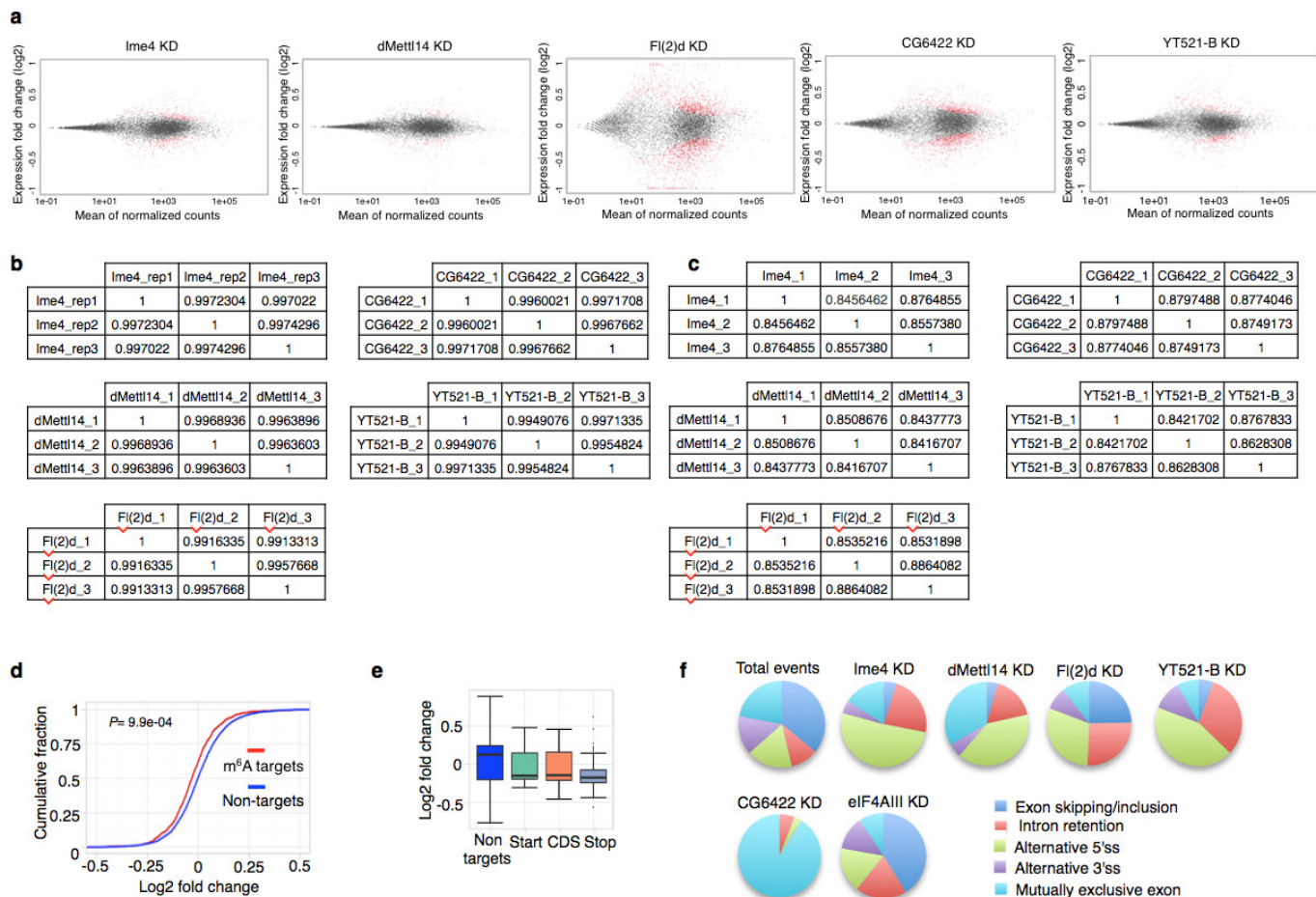
Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Characterization of the *Drosophila* m⁶A methyltransferase complex. **a**, Calibration curve for m⁶A nucleoside versus stable isotope labelled internal standard from digested *Escherichia coli* RNA. Areas under the curve (AUCs) are taken from MS/MS chromatograms. Amounts of 1–500 fmol m⁶A were evaluated. **b**, Calibration curve for external calibration of adenosine (A) nucleoside from 2–500 pmol. AUCs are extracted from chromatograms generated by ultraviolet detection. **c**, Phylogenetic analysis of METTL3 homologues. Each *Drosophila* (*D.m*) sequence clusters with the corresponding human (*H.s*), *Danio rerio* (*D.r*) and fungal orthologue. Fungi probably lost ancestral versions of individual methyltransferases with these families, with *Schizosaccharomyces pombe* (*S.p*) keeping only one orthologue (METTL4) and *Saccharomyces cerevisiae* (*S.c*) keeping the two other (METTL3 and METTL14). METTL3, METTL14 and METTL4 orthologues are indicated in green, blue and purple, respectively. See Methods for details about the tree construction. **d**, Box plots of average expression (rpkm) for all genes expressed by at least 1 rpkm in different conditions. Red dots indicate the position of m⁶A components in comparison to other expressed genes. Bottom, relative expression of target genes upon different knockdowns (KDs). The mean \pm s.d. of three technical measurements from three biological replicates is shown. **e**, Relative *vir* mRNA expression and levels of m⁶A in mRNA during *Drosophila* development. Number of hours post-fertilization for

different embryo, larval and pupal stages is indicated on the *x* axis. *vir* expression correlates with m⁶A levels. The mean \pm s.d. of three technical measurements from three biological replicates is shown. **f**, LC-MS/MS quantification of m⁶A levels in either control samples or in mRNA extracts depleted for the indicated proteins. *Vir* depletion affects m⁶A levels to the same extent as Fl(2)d knockdown. The mean \pm s.d. of three technical measurements from three biological replicates is shown for *fl(2)d* and mean \pm s.d. of three technical measurements from two biological replicates for *vir*. **g**, Co-immunoprecipitation of Myc–*Vir* with HA–Ime4 and HA–Fl(2)d. Extracts from S2R+ cells expressing HA-tagged proteins either with Myc alone or with Myc–*Vir* were immunoprecipitated using Myc-specific beads. Expression of indicated proteins was monitored by western blot analysis using anti-Myc and anti-HA antibodies. RNaseT1 treatment before immunoprecipitation is indicated at the bottom. **h**, Co-immunoprecipitation studies were carried out with lysates prepared from S2R+ cells co-expressing Myc–dMettl14 and HA–Ime4 upon control (Ctr) or Fl(2)d knockdown. For control experiments, S2R+ cells were transfected with Myc alone and HA–Ime4. Lysates were immunoprecipitated using anti-Myc antibody and then detected with anti-Myc and anti-HA antibodies. Knockdown of Fl(2)d weakens the interaction between Ime4 and dMettl14. **i**, Western blots showing Ime4 and dMettl14 protein expression levels in extracts from indicated genotypes. Tubulin is used as a loading control.

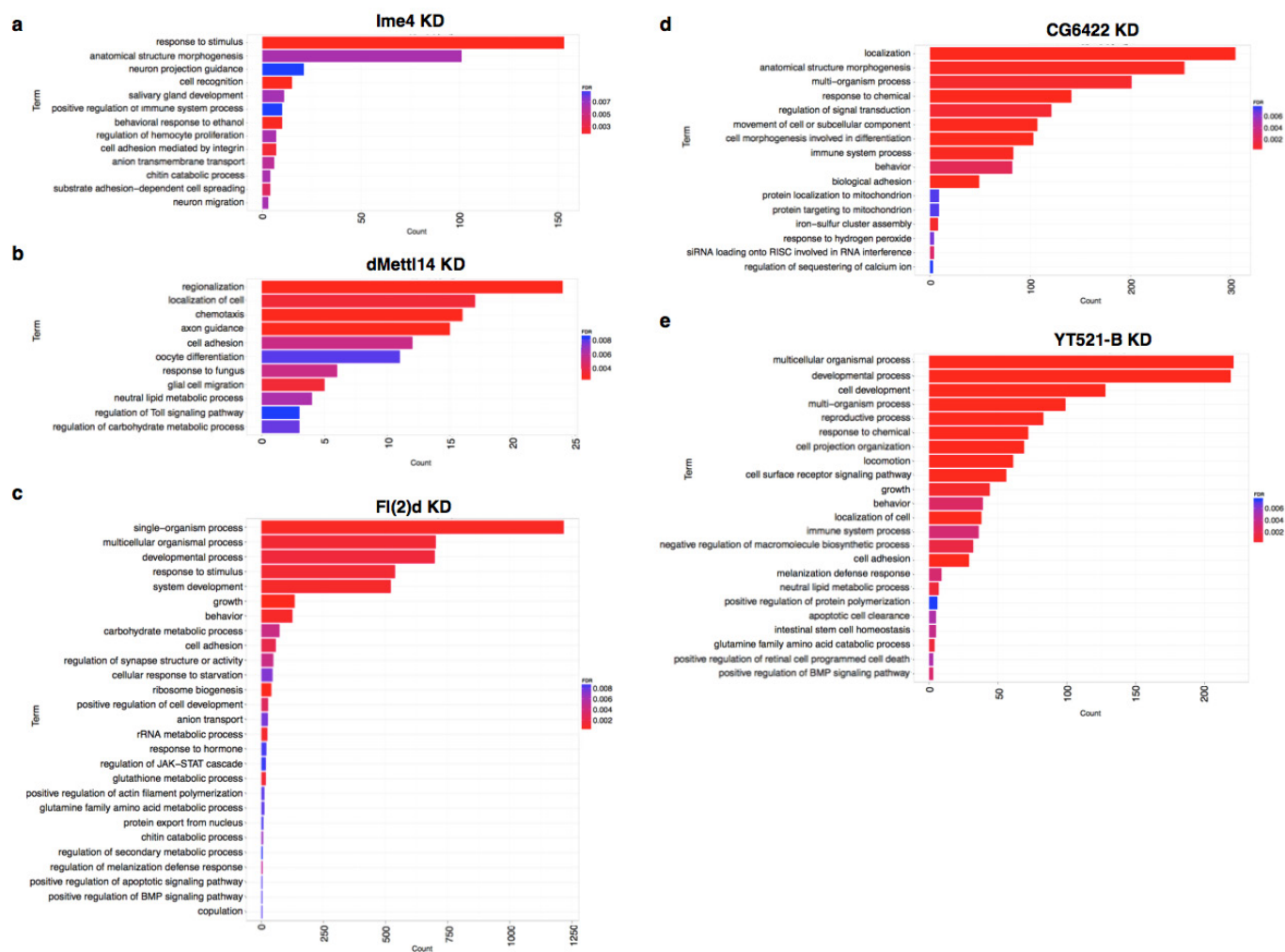


Extended Data Figure 2 | m⁶A quantification, MeRIP-seq validation and sequence features of m⁶A sites in *Drosophila* mRNA. **a**, Scatter plot of counts per million (CPM) values for intersected MeRIP peaks. The peaks have at least a support of 3 CPM in one of the replicates. **b**, qPCR validation of MeRIP peaks. Enrichment is calculated over a negative region in the *Rpl15* transcript. The mean ± s.d. of three technical measurements from two replicates is shown. **c**, Sequence motifs enriched in a fraction of m⁶A peaks, analysis performed by Homer.

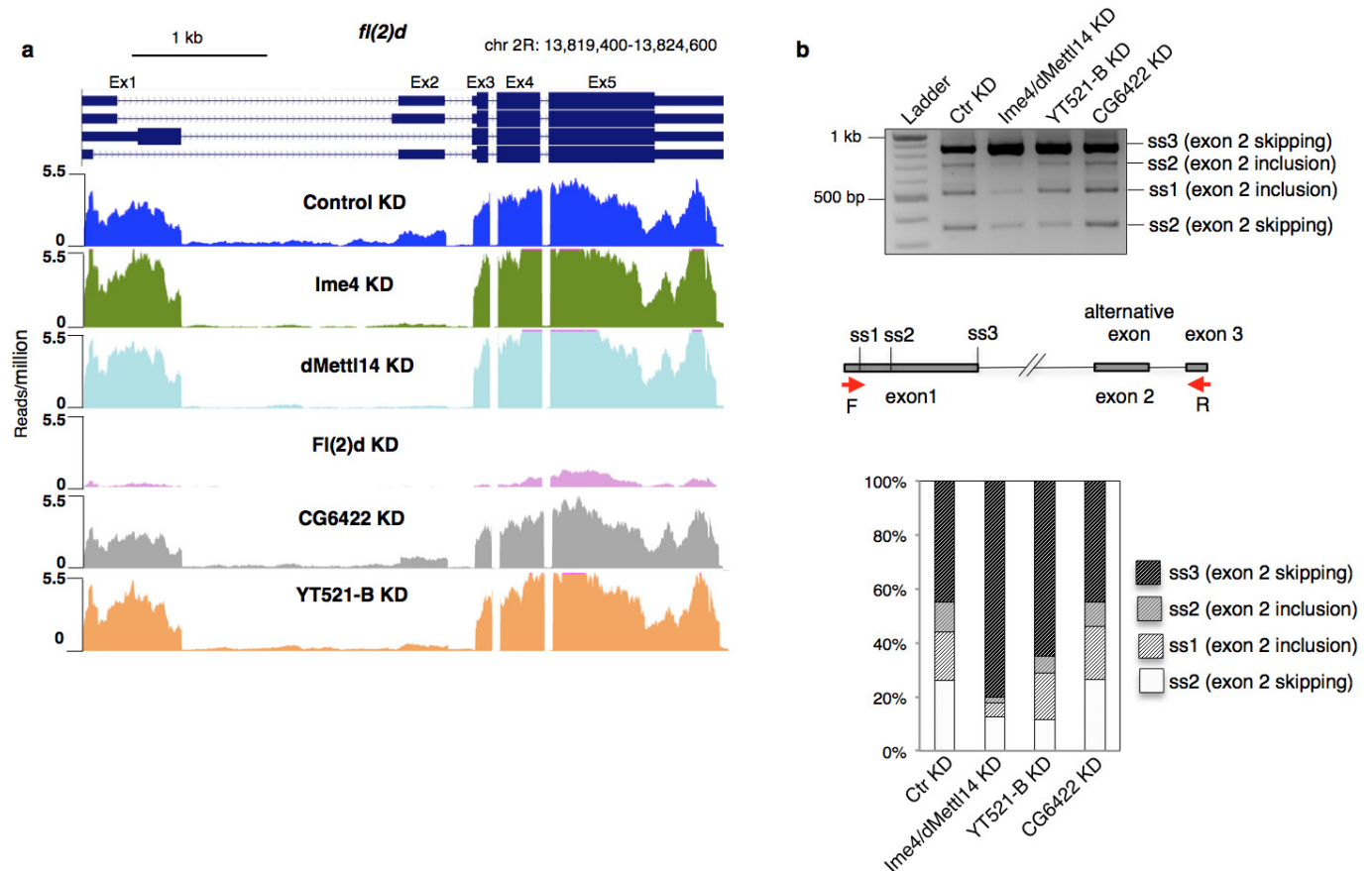


Extended Data Figure 3 | Significant fold changes and correspondence between biological replicates of RNA-seq data. **a**, Average versus mean-difference plots (MA-plots) show the moderated estimation of fold change and dispersion for RNA-seq data in the different knockdown conditions (adjusted P value < 0.05). The significant values are highlighted in red. **b**, Spearman sample-to-sample correlation based on gene expression profiles. **c**, Spearman sample-to-sample correlation based on splicing levels. **d**, Empirical cumulative distribution function (ECDF) plot of fold changes (\log_2) upon Ime4/dMettl14 double knockdown over control separated between m^6A targets and non-targets. Values are shown between -0.5 and 0.5 . The distributions were compared using Wilcoxon rank sum

test (P value $= 9.9 \times 10^{-4}$). **e**, Fold change upon Ime4/dMettl14 double knockdown versus control separated into genes without m^6A peaks (non-targets) or containing m^6A peaks within the CDS (CDS) or within a 300-bp window around the start or stop codon. Only genes considered for differential expression testing according to DESeq2 default filters are shown. **f**, Representation of differentially spliced events in the different knockdowns. Selection of 5' alternative splice sites and increase in intron retention are the two most enriched classes. Classification of splicing changes upon knockdown of the unrelated EJC component eIF4AIII is shown for comparison.

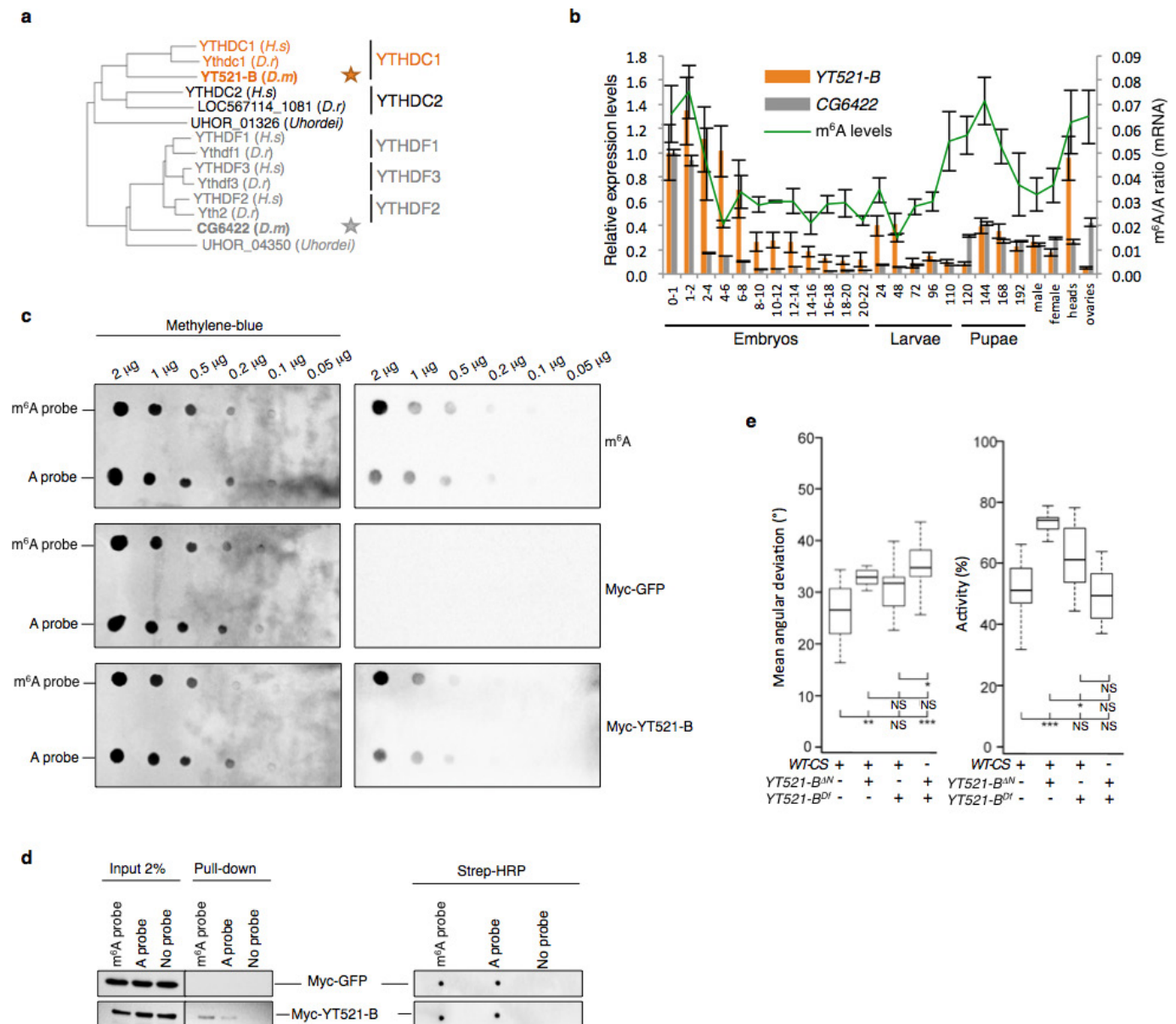


Extended Data Figure 4 | Gene ontology term enrichment analysis. a–e, Significant GO terms (adjusted P value < 0.05) of differentially expressed genes in Ime4 knockdown (a), dMettl14 knockdown (b), Fl(2)d knockdown (c), CG6422 knockdown (d) and YT521-B knockdown (e) cells versus control S2R⁺ cells. Analysis was performed using the Bioconductor package of GOSTats.



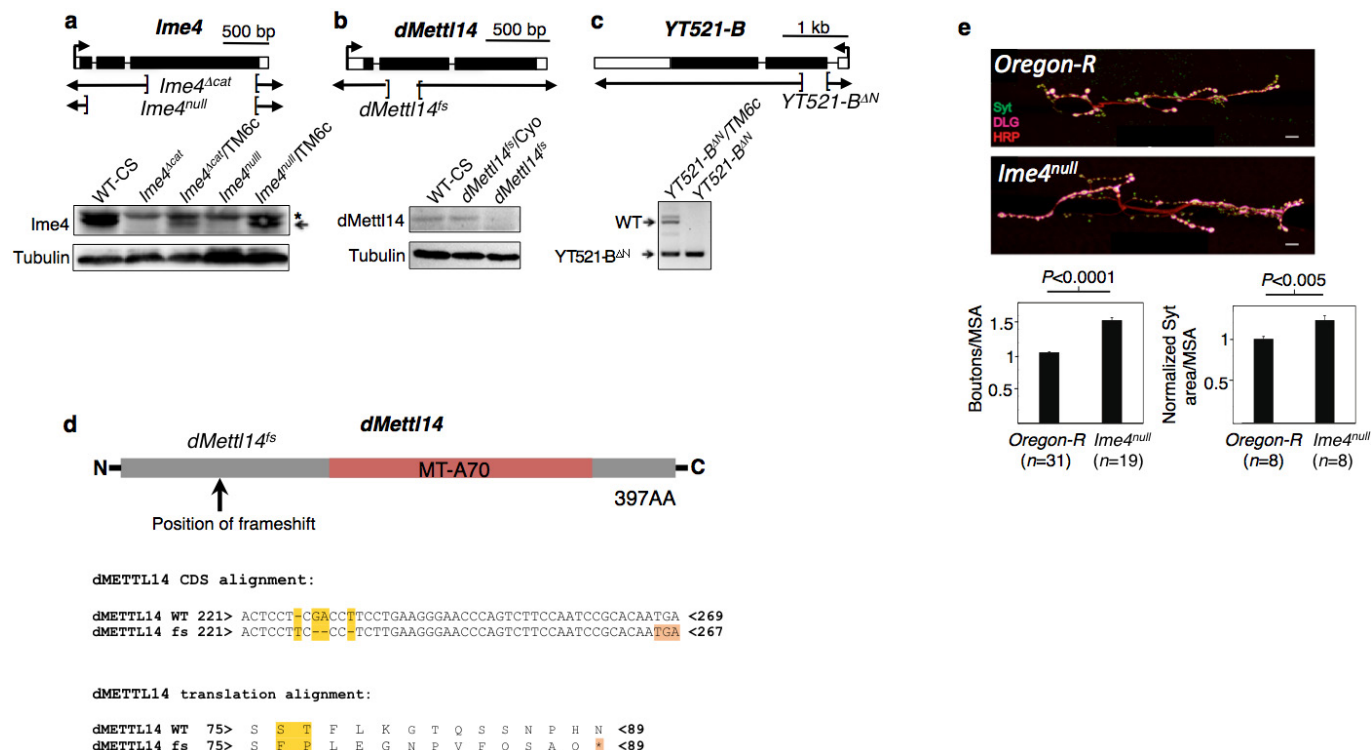
Extended Data Figure 5 | m⁶A nuclear components control *fl(2)d* splicing. **a**, UCSC Genome Browser screenshots of *fl(2)d* showing normalized RNA-seq data from control and indicated knockdown samples in S2R+ cells. The gene architecture of *fl(2)d* is shown at the top, with thin blue boxes representing the 5' and 3' UTRs, thick blue boxes representing the CDS, and thin lines representing introns. Exon numbers are indicated

at the top. Signals are displayed as RPM. **b**, Usage of different 5' splice sites in exon 1 of *fl(2)d* transcript and skipping of exon 2 upon different knockdowns. Analysis by semi-quantitative RT-PCR using primers in exon 1 and 3 (red arrows in the scheme). Quantification is indicated underneath the gel. ss1, splice site 1; ss2, splice site 2; ss3, splice site 3.



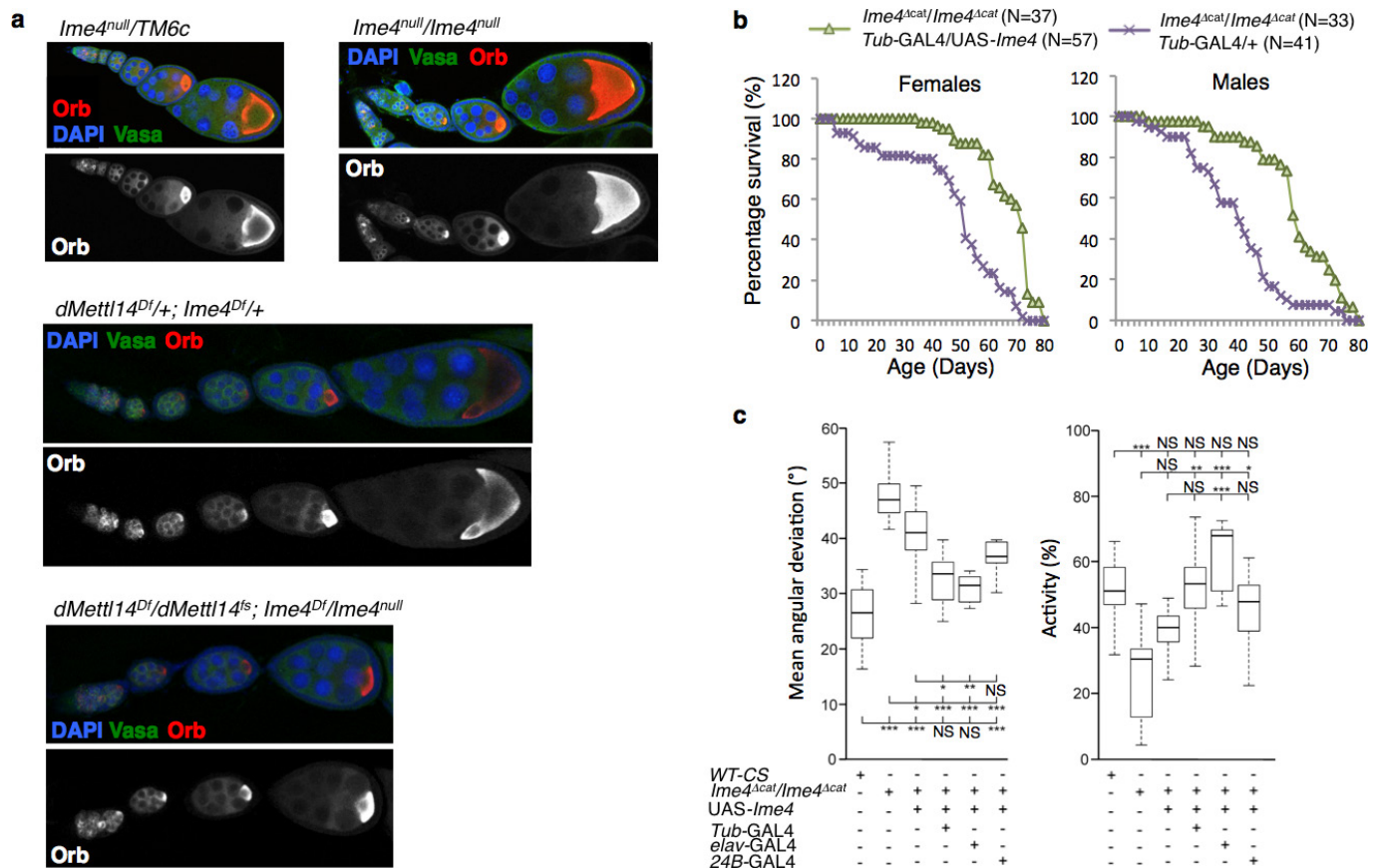
Extended Data Figure 6 | Characterization of *Drosophila* YTH components. **a**, Phylogenetic analysis. Sequences from *Ustilago hordei* (a basidiomycota fungi) were used, in the absence of appropriate sequences from *S. cerevisiae*, and worked as outliers for each cluster to show the separation between the two major groups. **b**, Relative expression of *YT521-B* and *CG6422* transcripts and levels of m^6A in mRNA during *Drosophila* development. Number of hours post-fertilization for different embryo, larval and pupal stages is indicated on the x axis. The mean \pm s.d. of three technical measurements from three biological replicates is shown. **c**, Dot-blot assay using biotinylated probe from *prolactin* transcripts with or without m^6A RNA modification. Protein extracts from S2R+ cells transfected with either Myc-GFP or Myc-YT521-B were analysed for binding specificity to the crosslinked probes. Left, methylene-blue staining of crosslinked probes. Right, immunostaining using anti-Myc or anti- m^6A

antibody. *YT521-B* protein shows the same enrichment to the methylated probe as anti- m^6A antibody. **d**, Pull-down using biotinylated m^6A probe from *prolactin* transcripts and protein extracts from S2R+ cells transfected with either Myc-GFP or Myc-YT521-B. The same probe lacking the methylation was used as a negative control. Left, western blot using anti-Myc antibody. Right, dot blot using anti-Strep-HRP antibody. The binding of Myc-YT521-B is increased with the methylated probe. Three independent experiments show similar results. **e**, Walking behaviour in Buridan's paradigm in heterozygous and transheterozygous *YT521-B* mutants. Left, median angular displacements from the direct approach to one of the stripes. Right, median fraction of time spent walking during a 15 min test period (Kruskal-Wallis analysis of variance with a Bonferroni correction). Fifteen female flies per genotype were used in both assays. NS, not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.



Extended Data Figure 7 | Genetic characterization of *Ime4*, *dMettl14* and *YT521-B*. **a–c**, Top, *Ime4* (**a**), *dMettl14* (**b**) and *YT521-B* (**c**) loci with indicated deletions. Bottom, loss of function for *Ime4* and *dMettl14* were monitored by western blot using respective endogenous antibodies, while anti-Tubulin antibody was used as a loading control. To analyse *YT521-B* deletion, PCR using genomic DNA from heterozygous or homozygous *YT521-B*^{ΔN} mutant flies was loaded on agarose gel. **d**, Scheme of the dMettl14 protein showing the conserved MT-A70 domain. The frameshift position caused by the guide RNA-induced deletions and the molecular nature of the allele are indicated below. **e**, Representative confocal images of muscle-6/7 NMJ synapses of abdominal hemisegment

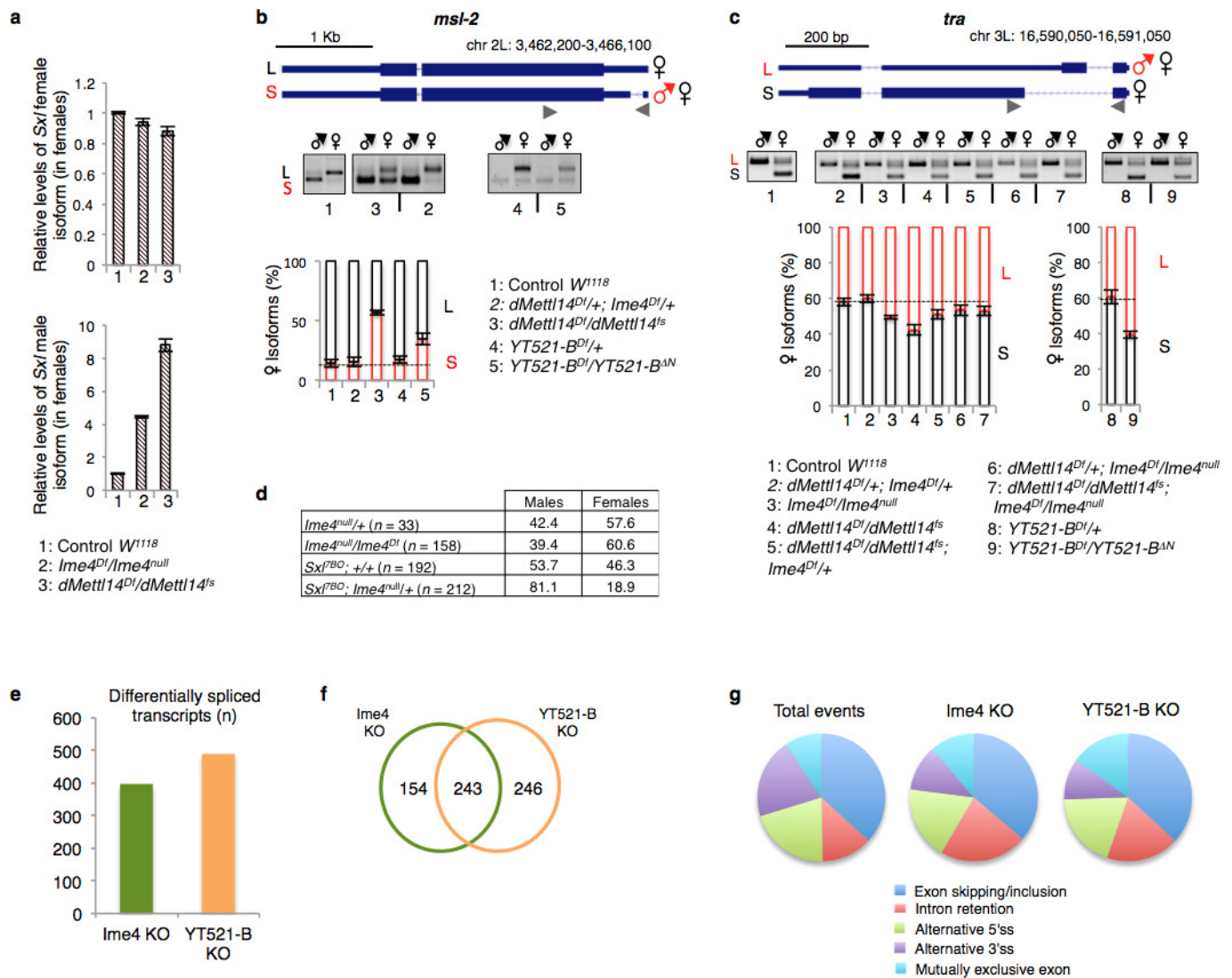
A2 for the indicated genotypes labelled with anti-DLG (magenta), anti-Synaptotagmin (green) and HRP (red) to reveal the synaptic vesicles and the neuronal membrane. Bottom, quantification of normalized bouton number (total number of boutons/muscle surface area ($\mu\text{m}^2 \times 1,000$)) and normalized Synaptotagmin area (total Synaptotagmin-positive area (μm^2)/muscle surface area ($\mu\text{m}^2 \times 1,000$)) of NMJ 6/7 in A3 of the indicated genotypes. Error bars show mean \pm s.e.m. *P* values were determined with a Student's *t*-test. The number of boutons and of active zones per boutons are increased upon *Ime4* knockout. MSA, muscle surface area.



Extended Data Figure 8 | *Ime4* mutant flies have reduced locomotion and shortened lifespan but apparent normal ovarian development.

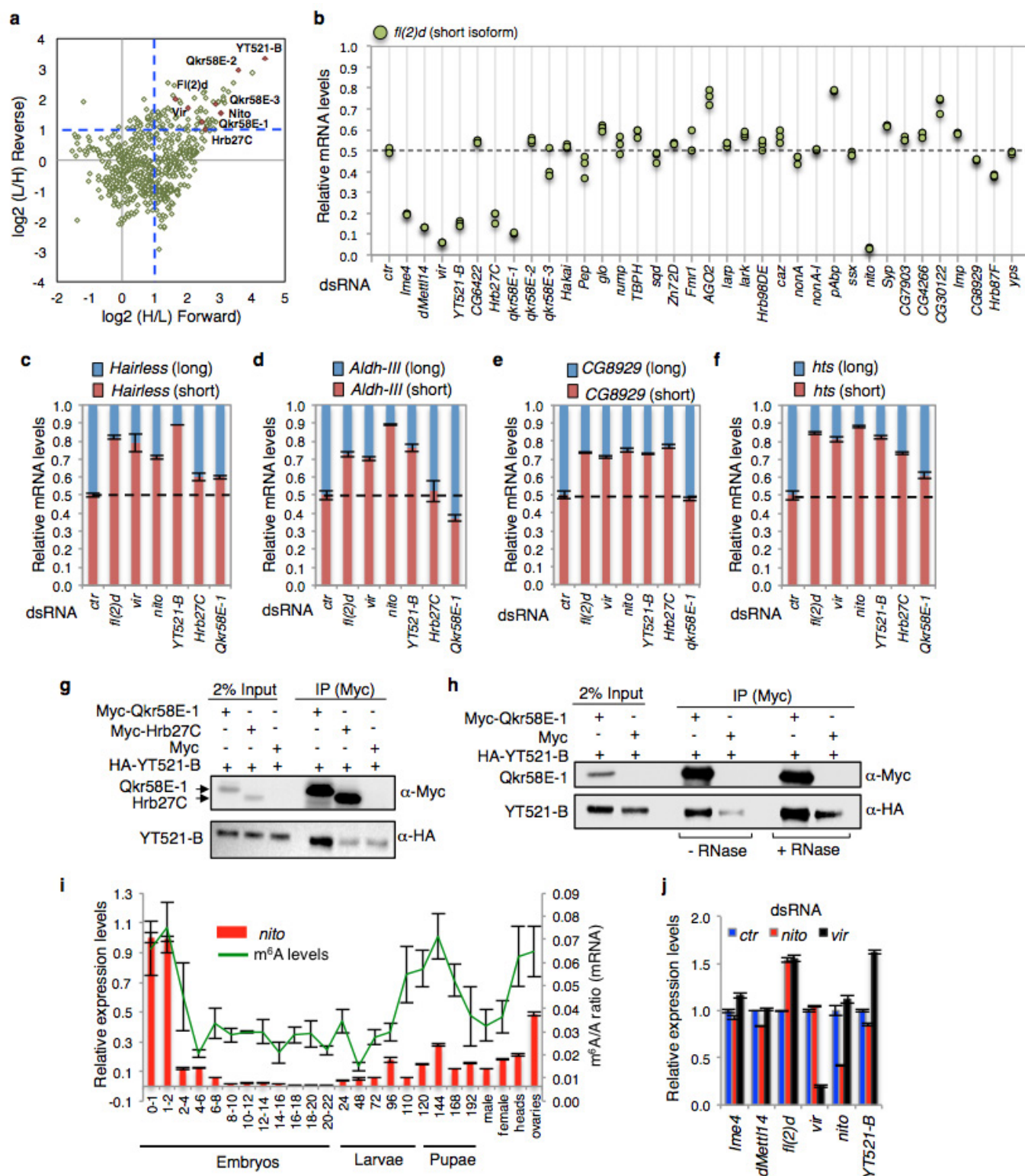
a, Ovarian immunostaining of indicated genotypes. DAPI (blue) stains nucleus, Vasa protein (Vasa) (green) shows the germ cells and Orb protein (Orb) (red) the oocyte. Only one oocyte per egg chamber is seen in control and mutant ovaries, arguing against encapsulation defects. **b**, Survival curves of adult *Drosophila*. The lifespan of *Ime4^{Δcat}* mutant flies (purple) and *Ime4^{Δcat}* mutant flies expressing *Ime4* cDNA ubiquitously (green) were

quantified for both females and males. **c**, Walking behaviour in Buridan's paradigm in *Ime4^{Δcat}* mutant flies or *Ime4^{Δcat}* mutant flies expressing *Ime4* cDNA ubiquitously (*Tub-GAL4*), in neurons (*elav-GAL4*) or in muscles (*24B-GAL4*). Left, median angular displacements from the direct approach to one of the stripes. Right, median fraction of time spent walking during a 15 min test period. Fifteen female flies per genotype. NS, not significant; **P* < 0.05, ***P* < 0.01, ****P* < 0.001 (Kruskal–Wallis analysis of variance with a Bonferroni correction).



Extended Data Figure 9 | m⁶A components fine-tune the sex determination pathway via YT521-B. **a**, Quantification of RT-qPCR experiments from RNA extracts of whole females using primers spanning exons 2 and 4 (top), as well as exons 2 and 3 (bottom) to quantify the levels of the *Sxl* female and male isoforms, respectively. The mean \pm s.d. of three technical measurements from two biological replicates is shown. **b**, Top, *msl-2* genome architecture with thin blue boxes representing the 5' and 3' UTRs, thick blue boxes representing the CDSs and thin lines representing introns. Arrowheads display the position of primers used for quantification. Bottom, spliced isoforms for *msl-2* were monitored by RT-PCR and PCR extracts were loaded on agarose gel. The quantification of three biological replicates is shown below as mean \pm s.d. **c**, Top, *tra* genome architecture with thin blue boxes representing the 5' and 3' UTRs,

thick blue boxes representing the CDSs and thin lines representing introns. Arrowheads display the position of primers used for quantification. Bottom, spliced isoforms for *tra* were monitored by RT-PCR and PCR extracts were loaded on agarose gels. The quantification of three biological replicates is shown below as mean \pm s.d. L, long isoform; S, short isoform. **d**, Table indicating the percentage of males and females hatching for the indicated genotypes. *Ime4* interacts genetically with *Sxl* to control female survival. **e**, Bar chart showing the number of differentially spliced genes upon knockout (KO) of *Ime4* and *YT521-B* in adult females. **f**, Venn diagram showing the overlap of targets in the indicated knockout. **g**, Pie charts showing distribution of splicing events in the different knockout conditions. Intron retention is overrepresented upon knockout of *Ime4* and *YT521-B* in vivo.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | RNA interference screen identifies Nito as a new member of the methyltransferase complex. **a**, SILAC-coupled mass spectrometry analysis using YT521-B-Myc as a bait. Scatterplot of normalized forward versus inverted reverse experiments plotted on a \log_2 scale. The threshold was set to a twofold enrichment (blue dashed line). Proteins in the top right quadrant are enriched in both duplicates. **b**, mRNA quantification of *fl(2)d* isoforms after knockdown of potential YT521-B-interacting proteins. Three proteins, Hrb27C, Qkr58E-1 and Nito, in addition to m⁶A components, control *fl(2)d* splicing in the same direction. Data points of three technical replicates are shown. **c–f**, mRNA quantification of m⁶A-regulated transcripts including *Hairless* (**a**), *Aldh-III* (**b**), CG8929 (**c**), *hts* (**d**) upon knockdown of indicated components. Nito controls m⁶A splicing events. The quantification of three technical replicates from two biological experiments is shown as mean \pm s.d. **g**, Co-immunoprecipitation studies were carried out with lysates prepared from S2R+ cells co-expressing Myc-Qkr58E-1, Myc-Hrb27C and HA-YT521-B. For control, S2R+ cells were transfected

with Myc alone and HA-YT521-B. Myc-containing proteins were immunoprecipitated using a Myc antibody and then immunoblotted with anti-Myc and anti-HA antibodies. **h**, Co-immunoprecipitation of Myc-Qkr58E-1 with HA-YT521-B with or without RNaseT1. Extracts from S2R+ cells expressing HA-YT521-B either with Myc control or with Myc-Qkr58E-1 were immunoprecipitated using Myc-specific beads. Expression of indicated proteins was monitored by immunoblotting using anti-Myc and anti-HA antibodies. **i**, Relative *nito* mRNA expression and levels of m⁶A in mRNA during *Drosophila* development. Number of hours post-fertilization for different embryo, larval and pupal stages is indicated on the x axis. *nito* expression correlates with m⁶A levels. The mean \pm s.d. of three technical measurements from three biological replicates are shown. **j**, Relative expression of indicated transcripts upon control, Nito and Vir knockdown. Vir and Nito knockdowns do not reduce expression of other components of the methyltransferase complex. The mean \pm s.d. of three technical measurements from two biological replicates is shown.

A massive, quiescent, population II galaxy at a redshift of 2.1

Mariska Kriek¹, Charlie Conroy², Pieter G. van Dokkum³, Alice E. Shapley⁴, Jieun Choi², Naveen A. Reddy⁵, Brian Siana⁵, Freeke van de Voort¹, Alison L. Coil⁶ & Bahram Mobasher⁵

Unlike spiral galaxies such as the Milky Way, the majority of the stars in massive elliptical galaxies were formed in a short period early in the history of the Universe. The duration of this formation period can be measured using the ratio of magnesium to iron abundance ($[\text{Mg}/\text{Fe}]$) in spectra^{1–4}, which reflects the relative enrichment by core-collapse and type Ia supernovae. For local galaxies, $[\text{Mg}/\text{Fe}]$ probes the combined formation history of all stars currently in the galaxy, including younger and metal-poor stars that were added during late-time mergers⁵. Therefore, to directly constrain the initial star-formation period, we must study galaxies at earlier epochs. The most distant galaxy for which $[\text{Mg}/\text{Fe}]$ had previously been measured⁶ is at a redshift of $z \approx 1.4$, with $[\text{Mg}/\text{Fe}] = 0.45^{+0.05}_{-0.19}$. A slightly earlier epoch ($z \approx 1.6$) was probed by combining the spectra of 24 massive quiescent galaxies, yielding an average $[\text{Mg}/\text{Fe}] = 0.31 \pm 0.12$ (ref. 7). However, the relatively low signal-to-noise ratio of the data and the use of index analysis techniques for both of these studies resulted in measurement errors that are too large to allow us to form strong conclusions. Deeper spectra at even earlier epochs in combination with analysis techniques based on full spectral fitting are required to precisely measure the abundance pattern shortly after the major star-forming phase ($z > 2$). Here we report a measurement of $[\text{Mg}/\text{Fe}]$ for a massive quiescent galaxy at a redshift of $z = 2.1$, when the Universe was three billion years old. With $[\text{Mg}/\text{Fe}] = 0.59 \pm 0.11$, this galaxy is the most Mg-enhanced massive galaxy found so far, having twice the Mg enhancement of similar-mass galaxies today. The abundance pattern of the galaxy is consistent with enrichment exclusively by core-collapse supernovae and with a star-formation timescale of 0.1 to 0.5 billion years—characteristics that are similar to population II stars in the Milky Way. With an average past star-formation rate of 600 to 3,000 solar masses per year, this galaxy was among the most vigorous star-forming galaxies in the Universe.

We observed the galaxy COSMOS-11494 with the near-infrared multi-object spectrograph MOSFIRE on the Keck I telescope⁸. It was also observed by two other programmes^{9,10}, and so we incorporated these publicly available archival data. COSMOS-11494 was selected from the 3D-HST survey^{11,12}. With a stellar mass M given by $\log_{10}(M/M_{\odot}) = 11.5 \pm 0.1$ (where M_{\odot} is the mass of the Sun and the errors given here and elsewhere represent 1σ), COSMOS-11494 is among the most massive galaxies at its redshift, and it has a very low star-formation rate of less than $0.6 M_{\odot} \text{ yr}^{-1}$ (see Methods). Similarly to the typical massive, quiescent galaxy at this redshift, it is smaller than its local counterparts of the same mass, with an effective radius of 2.1 kpc (ref. 13). The MOSFIRE rest-frame optical spectrum, the multi-wavelength spectral energy distribution¹¹ and Hubble Space Telescope (HST) colour image¹¹ of COSMOS-11494 are shown in Fig. 1.

Here we measure the stellar abundance pattern of COSMOS-11494 from the MOSFIRE rest-frame optical spectrum with our absorption

line fitter (alf) code¹⁴ (see Methods). For our default model we adopt a two-component stellar population, for which the age of both components and the slope of the stellar initial mass function (IMF) are free parameters. To enable comparison with previous work¹⁵, we also fit the spectrum with a single-age model and a Kroupa IMF¹⁶.

For the default model we find $[\text{Fe}/\text{H}] = -0.25 \pm 0.11$, $[\text{Mg}/\text{Fe}] = 0.59 \pm 0.11$, $[\text{Ca}/\text{Fe}] = 0.59 \pm 0.07$ and an age of 2.71 ± 0.22 Gyr; here $[X/Y] = \log_{10}(N_X/N_Y) - \log_{10}(N_{X,\odot}/N_{Y,\odot})$, where N_i is the atom number density of element i in COSMOS-11494 and $N_{i,\odot}$ is the corresponding

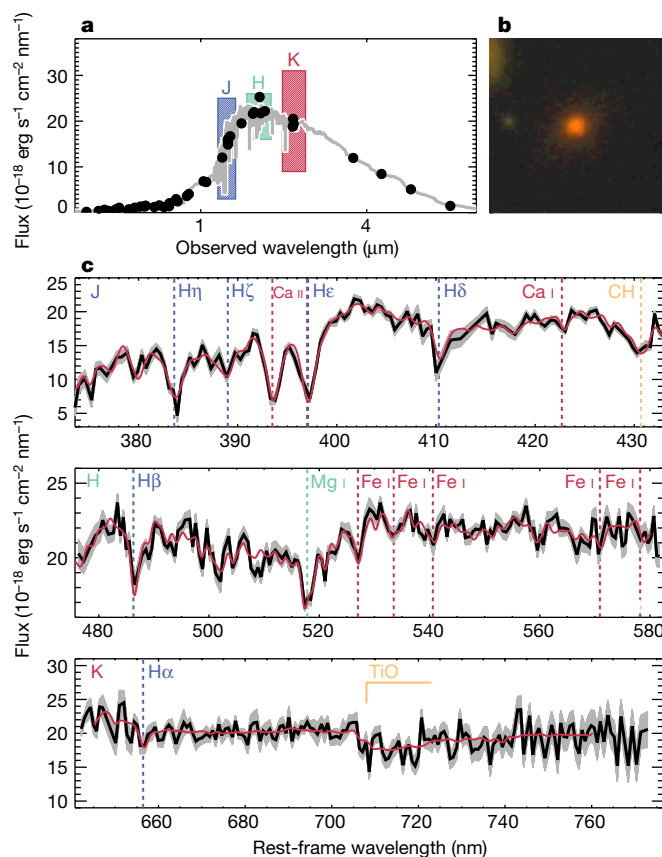


Figure 1 | Photometry, image and MOSFIRE spectrum of COSMOS-11494. **a**, Multi-wavelength spectral energy distribution (black circles) and best-fitting stellar population model to only the photometry (grey line). **b**, HST colour (V, J and H) image. **c**, MOSFIRE spectrum in three wavelength intervals (J, H and K; black), corresponding to the coloured areas in **a**. The grey shaded regions represent the 1σ uncertainty on the flux. The best-fitting stellar population model used to derive the age and abundance patterns is shown in red.

¹Department of Astronomy, University of California, Berkeley, California 94720, USA. ²Department of Astronomy, Harvard University, Cambridge, Massachusetts, USA. ³Astronomy Department, Yale University, New Haven, Connecticut, USA. ⁴Department of Physics and Astronomy, University of California, Los Angeles, California 90095, USA. ⁵Department of Physics and Astronomy, University of California, Riverside, California 92521, USA. ⁶Center for Astrophysics and Space Sciences, University of California, San Diego, La Jolla, California 92093, USA.

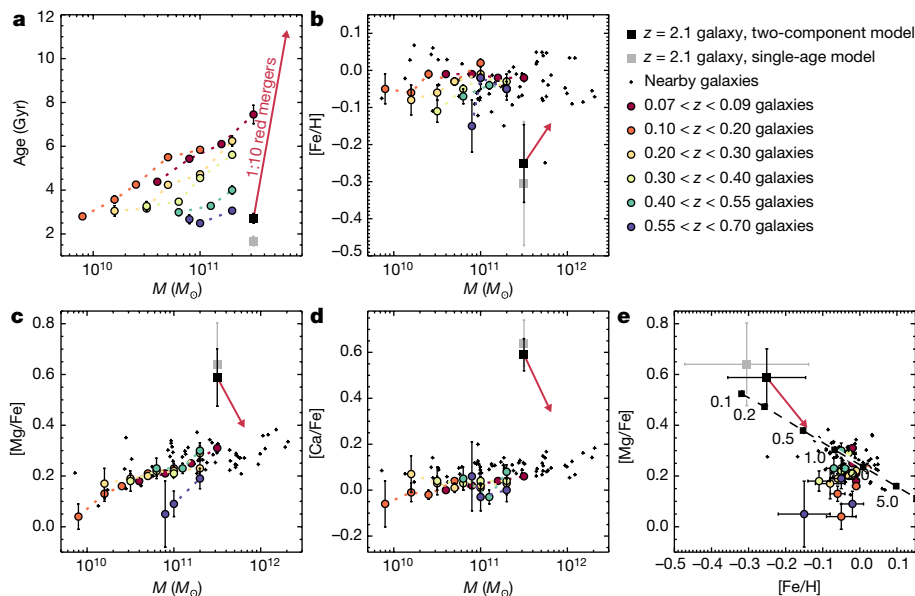


Figure 2 | Age and abundance patterns of COSMOS-11494 in comparison to lower-redshift quiescent galaxies. **a–d**, Age (**a**), $[\text{Fe}/\text{H}]$ (**b**), $[\text{Mg}/\text{Fe}]$ (**c**) and $[\text{Ca}/\text{Fe}]$ (**d**) of the stellar population versus stellar mass. **e**, $[\text{Mg}/\text{Fe}]$ versus $[\text{Fe}/\text{H}]$. The black dashed line represents a chemical evolution model for different star-formation timescales (indicated by the labels, in billions of years). In all panels, the black and

grey filled squares represent COSMOS-11494 for the two-component and single-age model, respectively, the coloured symbols represent low-redshift galaxies¹⁵ binned by mass and redshift, and the small black pluses are nearby galaxies¹⁷. Error bars are 1σ . The red arrows represent the simple evolutionary model (see main text).

value for the Sun. The best-fitting mass-to-light ratio (M/L) is consistent with the mass-to-light ratio assuming a Kroupa IMF (M_K/L_K ; $(M/L)/(M_K/L_K) = 0.97 \pm 0.55$), although the error is large because of the insufficient signal-to-noise ratio of the spectrum and the lack of rest-frame near-infrared coverage. We also fit this model with rest-frame wavelengths of less than $4,000 \text{ \AA}$ excluded, and find similar values. For the single-age model we find similar abundance ratios as for the two-component model, but the modelled age is 1 Gyr younger. This difference is expected, because younger stellar populations have lower mass-to-light ratios and so have larger weights in the fit.

In Fig. 2 we compare the spectral modelling results of COSMOS-11494 with those of galaxies at $0.05 < z < 0.7$ (ref. 15) and of a sample of nearby massive galaxies¹⁷. All galaxies are fitted with the alf code. Figure 2 illustrates that COSMOS-11494 is more Mg-enhanced than similar-mass galaxies at lower redshift, with $[\text{Mg}/\text{Fe}]$ about 0.3 dex higher. $[\text{Ca}/\text{Fe}]$ is also higher compared to the values for lower-redshift massive galaxies.

To interpret the abundance pattern of COSMOS-11494, we show a chemical evolution model in Fig. 2e, which assumes a Salpeter IMF¹⁸, a constant star-formation history over a given timescale, a core-collapse¹⁹ and a type Ia supernova yield model²⁰; we also adopt a power-law delay-time distribution of the form t^{-1} for type Ia supernovae that occurred between 0.1 Gyr and 13 Gyr (ref. 21). The star-formation timescale decreases along the curve, with the highest value of $[\text{Mg}/\text{Fe}]$ corresponding to the shortest timescale of 0.1 Gyr. The relatively low Fe abundance in combination with the high $[\text{Mg}/\text{Fe}]$ favours a short star-formation timescale of around 0.2 Gyr. Therefore, this model implies that COSMOS-11494 has experienced very little enrichment by type Ia supernovae.

However, the best-fitting timescale strongly depends on the assumed delay time of prompt type Ia supernovae. This parameter is poorly constrained in models and depends on the type Ia progenitor model²²; for the double degenerate scenario the lifetime can be as short as about 0.1 Gyr, whereas for a single degenerate scenario it can be as high as about 0.5 Gyr (ref. 22). However, the delay times of prompt type Ia supernovae inferred from observations are as short as 0.1 Gyr (ref. 21), and so 0.5 Gyr may be a conservative upper limit. The assumed IMF affects the chemical evolution model as well, and a flatter IMF

results in a longer timescale. Finally, the adopted chemical evolution model depends on the core-collapse-supernova yields¹⁹, and other yield models were unable to reproduce the observed high $[\text{Mg}/\text{Fe}]$ in combination with the observed $[\text{Fe}/\text{H}]$ ^{23,24}. Nonetheless, individual stars with similarly high $[\text{Mg}/\text{Fe}]$ and $[\text{Fe}/\text{H}]$ as COSMOS-11494 have been identified in the bulge of the Milky Way²⁵, which supports the validity of the adopted yield model¹⁹. Taking into account all uncertainties on our chemical evolution model, we estimate a star-formation timescale of 0.1–0.5 Gyr.

Ca, which is also produced and returned to the interstellar medium through core-collapse supernovae, also shows a strong enhancement with respect to Fe; the difference compared to low-redshift analogues is even more extreme than for Mg. This differential evolution of Ca and Mg with time is unexpected, because both elements are formed in massive stars, but metallicity-dependent core-collapse-supernova yields might explain the differences¹⁹.

When combining the star-formation timescale and the best-fitting stellar mass, we find an average past star-formation rate of $(600\text{--}3,000) M_\odot \text{ yr}^{-1}$. The single-age best-fitting stellar population synthesis (SPS) model sets a lower limit on the formation redshift of $z > 4$. For the two-component model we find an average formation redshift of $z = 12_{-4}^{+9}$. The inferred star-formation rate and formation epoch are consistent with the properties that were derived for the most active galaxy found so far, HFLS3²⁶. This dusty sub-millimetre galaxy has a star-formation rate of $2,900 M_\odot \text{ yr}^{-1}$ and a redshift of $z = 6.34$, and so could be similar to the star-forming progenitor of COSMOS-11494.

COSMOS-11494 seems to be more Mg-enhanced than $z \approx 1.5$ galaxies in previous work^{6,7}. However, because different methods have been used in these studies, systematic differences may occur, and the derived values cannot be directly compared⁴. For a more direct comparison to previous studies, we use the Lick indices²⁷ $\langle \text{Fe} \rangle = (\text{Fe}5270 + \text{Fe}5335)/2$ and $\text{Mg } b$, where Fe5270 and Fe5335 are sensitive to the Fe lines near wavelengths of $5,270 \text{ \AA}$ and $5,335 \text{ \AA}$, respectively, and $\text{Mg } b$ is sensitive to the Mg I triplet at $5,165\text{--}5,185 \text{ \AA}$. For the $z = 1.4$ galaxy, the spectrum was not deep enough to measure the two Fe lines needed to determine $\langle \text{Fe} \rangle$. Instead, we calibrated $\langle \text{Fe} \rangle$ for two other Fe lines (Fe4388 and Fe5015) using a set of SPS models²⁸, and derived $\langle \text{Fe} \rangle$ from these (marginally detected) lines for the $z = 1.4$

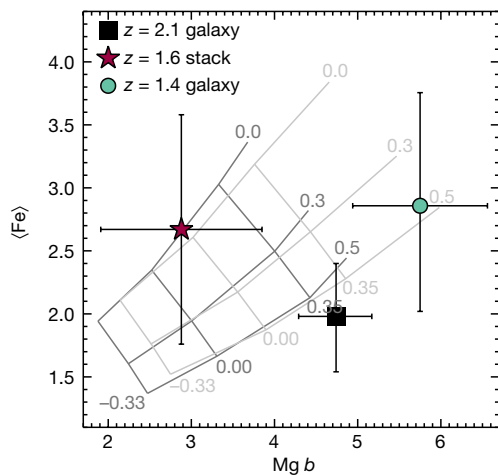


Figure 3 | Abundance pattern of COSMOS-11494 in comparison to other high-redshift galaxies. We show the Lick indices $\langle \text{Fe} \rangle$ and $\text{Mg } b$ for COSMOS-11494 (black square), a $z \approx 1.4$ quiescent galaxy⁶ observed with VLT/X-Shooter (blue circle), and a stack of $z \approx 1.6$ quiescent galaxy spectra⁷ observed with Subaru/MOIRCS (pink star). Error bars are 1σ . We also show grids based on SPS models²⁸ for a range of values of $[\text{Z}/\text{H}]$ (-0.33 , 0.00 and 0.35) and $[\text{Mg}/\text{Fe}]$ (0.0 , 0.3 and 0.5). The dark and light grey grids are for galaxy ages of 2 Gyr and 4 Gyr, respectively.

galaxy. Figure 3 shows the measurements in comparison to a grid of SPS models²⁸ for a range of metallicities ($[\text{Z}/\text{H}] = -0.33$, 0.00 and 0.35) and $[\text{Mg}/\text{Fe}]$ values (0.0 , 0.3 and 0.5).

For COSMOS-11494, the $[\text{Mg}/\text{Fe}] \approx 0.6$ implied from the Lick indices is consistent with the modelling results. This value illustrates that COSMOS-11494—which probes an earlier epoch than previous work^{6,7}—is indeed more Mg-enhanced than the quiescent galaxies in the other studies. Furthermore, our high signal-to-noise spectrum results in the most robust abundance pattern measurement for a distant galaxy so far. Although the two different approaches to derive the abundance pattern give consistent results for COSMOS-11494, for distant galaxies, the full spectral modelling approach is strongly preferred over the approach involving Lick indices (Methods).

The high $[\text{Mg}/\text{Fe}]$ of COSMOS-11494 compared to lower-redshift quiescent galaxies of similar mass suggests that this galaxy, and possibly other distant quiescent galaxies, do not passively evolve into quiescent early-type galaxies today. A similar conclusion was drawn from the small sizes of distant quiescent galaxies compared to their local analogues²⁹, and in the past several years it has become apparent that distant quiescent galaxies grow in mass and size by accreting primarily smaller galaxies^{5,30}. This inside-out growth by late-time mergers with less-massive galaxies predicts a decline in $[\text{Mg}/\text{Fe}]$, because lower-mass galaxies are less Mg-enhanced^{3,15}.

In Fig. 2 we explore this scenario by showing the predicted path of COSMOS-11494 for a simple evolutionary model. We assume that the galaxy grows by red minor (1:10) mergers with smaller, less-Mg-enhanced quiescent galaxies with $[\text{Mg}/\text{Fe}] = 0.18$, $[\text{Fe}/\text{H}] = -0.05$ and $[\text{Ca}/\text{Fe}] = 0.02$ (ref. 15), and that the mass nearly doubles between $z = 2.1$ and the present day⁵ following the mass evolution $d[\log(M)]/dz = -0.15$. Therefore, we assume no evolution in $[\text{Mg}/\text{Fe}]$ and $[\text{Ca}/\text{Fe}]$ at lower masses. If the abundance ratios for these galaxies would be higher at earlier times as well, then the predicted evolution would be less strong. To estimate the evolution in galaxy age for the merger model, we assume that the age is proportional to $M^{0.3}$, as was found for $z < 0.7$ galaxies¹⁵.

Figure 2 shows that the merger model can substantially decrease $[\text{Mg}/\text{Fe}]$ and $[\text{Ca}/\text{Fe}]$, and increase $[\text{Fe}/\text{H}]$. However, there are several caveats to our simple model comparison. First, the $0.07 < z < 0.7$ measurements are derived by fitting a single-age model, and so are sensitive to low levels of recent star formation. Second, we assume that

the accreted stars are well mixed with the *in situ* population. However, simulations of galaxy formation show that the added material is mostly deposited in the outskirts of the galaxy³⁰ and so the net evolution due to mergers—in the central parts targeted by the spectrographs—may be less. Third, it is unlikely that the descendant of COSMOS-11494 is a typical massive, quiescent galaxy today. The star formation in many low-redshift quiescent galaxies is quenched at later times, resulting in longer star-forming periods and, hence, lower $[\text{Mg}/\text{Fe}]$. Therefore, the descendant of COSMOS-11494 presumably resides in the tail of the low-redshift distributions. Finally, the model does not include possible late-time star formation or mergers with star-forming galaxies, which would also result in a decrease in $[\text{Mg}/\text{Fe}]$ with time.

More spectra of quiescent galaxies at high redshifts are needed to measure the evolution of the slope and the intercept of the age– M and $[\text{Mg}/\text{Fe}]$ – M relations. Such measurements could eventually discriminate between different evolutionary scenarios, and the amount of mixing of stars after galaxy mergers¹⁵. In combination with more accurate supernova progenitor and yield models, and therefore improved chemical evolution models, these measurements will also provide unique information on the star-formation histories of the most massive galaxies and their possible role in the reionization of the Universe at $z > 6$. We expect that observations with NIRSpec on the James Webb Space Telescope will revolutionize this field within the next five years, with future ultra-deep observations with MOSFIRE paving the way.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 May; accepted 17 October 2016.

- Matteucci, F. Abundance ratios in ellipticals and galaxy formation. *Astron. Astrophys.* **288**, 57–64 (1994).
- Trager, S. C., Faber, S. M., Worthey, G. & González, J. J. The stellar population histories of early-type galaxies. II. Controlling parameters of the stellar populations. *Astron. J.* **120**, 165–188 (2000).
- Thomas, D., Maraston, C., Bender, R. & Mendes de Oliveira, C. The epochs of early-type galaxy formation as a function of environment. *Astrophys. J.* **621**, 673–694 (2005).
- Conroy, C., Graves, G. J. & van Dokkum, P. G. Early-type galaxy archeology: ages, abundance ratios, and effective temperatures from full-spectrum fitting. *Astrophys. J.* **780**, 33 (2014).
- van Dokkum, P. G. *et al.* The growth of massive galaxies since $z = 2$. *Astrophys. J.* **709**, 1018–1041 (2010).
- Loncne, I. *et al.* Old age and supersolar metallicity in a massive $z \sim 1.4$ early-type galaxy from VLT/X-Shooter spectroscopy. *Mon. Not. R. Astron. Soc.* **454**, 3912–3919 (2015).
- Onodera, M. *et al.* The ages, metallicities, and element abundance ratios of massive quenched galaxies at $z \approx 1.6$. *Astrophys. J.* **808**, 161 (2015).
- McLean, I. S. *et al.* MOSFIRE, the multi-object spectrometer for infra-red exploration at the Keck Observatory. *Proc. SPIE* **8446**, 84460J (2012).
- Belli, S., Newman, A. B., Ellis, R. S. & Konidaris, N. P. MOSFIRE absorption line spectroscopy of $z > 2$ quiescent galaxies: probing a period of rapid size growth. *Astrophys. J.* **788**, L29 (2014).
- Kriek, M. *et al.* The MOSFIRE Deep Evolution Field (MOSDEF) survey: rest-frame optical spectroscopy for ~ 1500 H -selected galaxies at $1.37 \leq z \leq 3.8$. *Astrophys. J. Suppl. Ser.* **218**, 15 (2015).
- Skelton, R. E. *et al.* 3D-HST WFC3-selected photometric catalogs in the five CANDELS/3D-HST fields: photometry, photometric redshifts, and stellar masses. *Astrophys. J. Suppl. Ser.* **214**, 24 (2014).
- Momcheva, I. G. *et al.* The 3D-HST survey: Hubble Space Telescope WFC3/G141 grism spectra, redshifts, and emission line measurements for $\sim 100,000$ galaxies. Preprint at <https://arxiv.org/abs/1510.02106> (2015).
- van de Sande, J. *et al.* Stellar kinematics of $z \sim 2$ galaxies and the inside-out growth of quiescent galaxies. *Astrophys. J.* **771**, 85 (2013).
- Conroy, C. & van Dokkum, P. Counting low-mass stars in integrated light. *Astrophys. J.* **747**, 69 (2012).
- Choi, J. *et al.* The assembly histories of quiescent galaxies since $z = 0.7$ from absorption line spectroscopy. *Astrophys. J.* **792**, 95 (2014).
- Kroupa, P. On the variation of the initial mass function. *Mon. Not. R. Astron. Soc.* **322**, 231–246 (2001).
- Conroy, C. & van Dokkum, P. G. The stellar initial mass function in early-type galaxies from absorption line spectroscopy. II. *Results. Astrophys. J.* **760**, 71 (2012).
- Salpeter, E. E. The luminosity function and stellar evolution. *Astrophys. J.* **121**, 161–167 (1955).

19. Kobayashi, C., Umeda, H., Nomoto, K., Tominaga, N. & Ohkubo, T. Galactic chemical evolution: carbon through zinc. *Astrophys. J.* **653**, 1145–1171 (2006).
20. Nomoto, K., Thielemann, F.-K. & Yokoi, K. Accreting white dwarf models of type I supernovae. III. Carbon deflagration supernovae. *Astrophys. J.* **286**, 644–658 (1984).
21. Maoz, D., Mannucci, F. & Brandt, T. D. The delay-time distribution of type Ia supernovae from Sloan II. *Mon. Not. R. Astron. Soc.* **426**, 3282–3294 (2012).
22. Kobayashi, C. & Nomoto, K. The role of type Ia supernovae in chemical evolution. I. Lifetime of type Ia supernovae and metallicity effect. *Astrophys. J.* **707**, 1466–1484 (2009).
23. Thielemann, F.-K., Nomoto, K. & Hashimoto, M.-A. Core-collapse supernovae and their ejecta. *Astrophys. J.* **460**, 408–436 (1996).
24. Woosley, S. E. & Weaver, T. A. The evolution and explosion of massive stars. II. Explosive hydrodynamics and nucleosynthesis. *Astrophys. J. Suppl. Ser.* **101**, 181–235 (1995).
25. Fulbright, J. P., McWilliam, A. & Rich, R. M. Abundances of Baade's window giants from Keck HIRES spectra. II. The alpha and light odd elements. *Astrophys. J.* **661**, 1152–1179 (2007).
26. Riechers, D. A. *et al.* A dust-obscured massive maximum-starburst galaxy at a redshift of 6.34. *Nature* **496**, 329–333 (2013).
27. Worthey, G., Faber, S. M., Gonzalez, J. J. & Burstein, D. Old stellar populations. V. Absorption feature indices for the complete LICK/IDS sample of stars. *Astrophys. J. Suppl. Ser.* **94**, 687–722 (1994).
28. Thomas, D., Maraston, C. & Bender, R. Stellar population models of Lick indices with variable element abundance ratios. *Mon. Not. R. Astron. Soc.* **339**, 897–911 (2003).
29. van Dokkum, P. G. *et al.* Confirmation of the remarkable compactness of massive quiescent galaxies at $z \sim 2.3$: early-type galaxies did not form in a simple monolithic collapse. *Astrophys. J.* **677**, L5–L8 (2008).
30. Naab, T., Johansson, P. H. & Ostriker, J. P. Minor mergers and the size evolution of elliptical galaxies. *Astrophys. J.* **699**, L178–L182 (2009).

Acknowledgements M.K. acknowledges discussions with J. Greene and E. Quataert. The data presented in this paper were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership between the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the financial support of the W. M. Keck Foundation. We acknowledge the cultural role that the summit of Mauna Kea has within the indigenous Hawaiian community. We acknowledge support from NSF AAG collaborative grants AST-1312780, 1312547, 1312764 and 1313171 and archival grant AR-13907, provided by NASA through a grant from the Space Telescope Science Institute. C.C. acknowledges support from NASA grant NNX13AI46G, NSF grant AST-1313280 and the Packard Foundation.

Author Contributions M.K., P.G.v.D. and C.C. wrote the primary Keck proposal. M.K. and C.C. led the interpretation. M.K. wrote the reduction pipeline, reduced the data, determined the stellar mass, measured the Lick indices and wrote the text. C.C. developed the SPS model, fitted the spectrum and derived the chemical evolution model. M.K., P.G.v.D., J.C., F.v.d.V. and N.A.R. did the observations. All authors contributed to the analysis and interpretation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.K. (mkriek@berkeley.edu).

Reviewer Information *Nature* thanks T. C. Beers, C. Kobayashi and C. Maraston for their contribution to the peer review of this work.

METHODS

Best-fitting model to the photometry. To derive the stellar mass of COSMOS-11494, we fit the broadband photometry with the flexible stellar population synthesis (SPS) models^{31,32}. We assume a delayed exponential star formation history of the form $\text{SFR} \propto te^{-t/\tau}$ and the dust attenuation law from ref. 33. We adopt the Chabrier stellar IMF³⁴ to facilitate direct comparison with lower-redshift studies¹⁵.

The error bar on the stellar mass is completely dominated by systematic uncertainties. We estimate this uncertainty by varying the SPS model^{35,36}, dust attenuation law^{37,38}, parameterization of the star-formation history, and the scaling of the broadband spectral energy distribution. We do not vary the IMF, because it can be approximated by a simple offset in the stellar mass. A Kroupa IMF¹⁶ would have resulted in a similar stellar mass as for the Chabrier IMF³⁴, whereas a Salpeter IMF¹⁸ would have resulted in a stellar mass a factor of 1.6 higher.

Spectral fitting. Parameters are estimated from the rest-frame optical spectrum with the `alf` code^{4,14,15}. This code combines libraries of isochrones and empirical stellar spectra with synthetic stellar spectra covering a wide range of elemental abundance patterns. The code fits for C, N, O, Na, Mg, Ca, Ti, V, Cr, Mn, Fe, Co, Ni, redshift, velocity dispersion and several emission lines. The stellar population age and IMF are free parameters as well, and multiple stellar population components are allowed. When fitting a two-component model, the age represents the mass-weighted average age of the two separate components. The ratio of the model and data are fitted by a high-order polynomial to avoid potential issues with the flux calibration of the data. The fitting is done using a Markov chain Monte Carlo algorithm³⁹.

Lick indices versus full spectral modelling. As mentioned in the main text, for COSMOS-11494 the $[\text{Mg}/\text{Fe}] \approx 0.6$ implied from the Lick indices is consistent with the modelling results. The metallicity measurements also agree between the two methods: our best-fitting values for $[\text{Mg}/\text{Fe}]$ and $[\text{Fe}/\text{H}]$ imply $[Z/\text{H}] = [\text{Fe}/\text{H}] + 0.94[\text{Mg}/\text{Fe}] = 0.25$ (ref. 28), which is consistent with the model shown in Fig. 3 for the best-fitting age of 2.5 Gyr. For the $z \approx 1.4$ individual galaxies and the $z \approx 1.6$ stack, the abundance patterns are based on the Lick indices, and so by construction they should be closer to the grid points. This is indeed the case for the $z \approx 1.4$ individual galaxy⁶, with $[\text{Mg}/\text{Fe}] = 0.45^{+0.05}_{-0.19}$ and $[Z/\text{H}] = 0.61^{+0.06}_{-0.05}$. For the derived age of 4 Gyr, the two Lick indices are consistent with the model grid. However, the individual Fe lines yield very different and inconsistent results when calculating $\langle \text{Fe} \rangle$. For the $z \approx 1.6$ galaxy stack⁷, the derived $[\text{Mg}/\text{Fe}] = 0.31^{+0.12}_{-0.12}$, $[Z/\text{H}] = 0.24^{+0.20}_{-0.14}$ and $\log(\text{age}/\text{Gyr}) = 0.04^{+0.10}_{-0.08}$ are less consistent with the derived values based on all Lick indices, although the error bars are large.

Although the two different approaches to derive the abundance pattern agree well for COSMOS-11494, the full spectral modelling approach is strongly preferred over the approach involving Lick indices. The rest-frame optical spectrum of $z > 1$ galaxies has been shifted to near-infrared wavelengths. The many skylines at these wavelengths in combination with the relatively lower signal-to-noise ratio of the spectra of distant galaxies result in large error bars on the measurement of a single feature. Lick indices are integrated measurements and do not take into account wavelength-dependent features within the bandpass. Therefore, skylines

will severely complicate their measurement because affected wavelengths are not down-weighted. Furthermore, skylines result in non-Gaussian and correlated noise properties, and so error bars on Lick indices are usually underestimated. Consequently, Lick indices are much more prone to systematic errors. The discrepancies between individual Lick indices and the derived stellar abundance pattern based on all Lick indices for the two lower-redshift measurements further illustrate this point. By modelling the full spectrum, we use many more features and can better deal with the larger uncertainties in regions affected by skylines.

Code availability. The data reduction package used to process the raw MOSFIRE data will be made public in the coming year at <http://astro.berkeley.edu/~mariska>. To derive the stellar mass, we used the Flexible SPS models that are available at <https://github.com/cconroy20/fps> and the SPS fitting code FAST⁴⁰, which is publicly available at <http://astro.berkeley.edu/~mariska/fast/>. The spectral fitting code `alf`^{4,14,15} that was used to derive the abundance pattern is not publicly available, but the underlying model components are available for download from <http://scholar.harvard.edu/cconroy/sps-models>.

Data availability. The one-dimensional original and binned spectrum shown in Fig. 1 and the corresponding 1σ uncertainties, as well as the best-fitting model spectrum are available as Source Data. The binned spectrum is constructed by first masking wavelengths affected by skylines and poor atmospheric transmission, and then taking the median of the flux of ten non-masked consecutive pixels. The photometric data points shown in Fig. 1 are made available by the 3D-HST collaboration (catalogue version v4.1) at <http://3dhst.research.yale.edu/Data.php>. The abundance pattern for COSMOS-11494 for the two-component and single-age models, as shown in Fig. 2, is also available as Source Data.

- Conroy, C., Gunn, J. E. & White, M. The propagation of uncertainties in stellar population synthesis modeling. I. The relevance of uncertain aspects of stellar evolution and the initial mass function to the derived physical properties of galaxies. *Astrophys. J.* **699**, 486–506 (2009).
- Conroy, C. & Gunn, J. E. The propagation of uncertainties in stellar population synthesis modeling. III. Model calibration, comparison, and evaluation. *Astrophys. J.* **712**, 833–857 (2010).
- Kriek, M. & Conroy, C. The dust attenuation law in distant galaxies: evidence for variation with spectral type. *Astrophys. J.* **775**, L16 (2013).
- Chabrier, G. Galactic stellar and substellar initial mass function. *Publ. Astron. Soc. Pacif.* **115**, 763–795 (2003).
- Bruzual, G. & Charlot, S. Stellar population synthesis at the resolution of 2003. *Mon. Not. R. Astron. Soc.* **344**, 1000–1028 (2003).
- Maraston, C. Evolutionary population synthesis: models, analysis of the ingredients and application to high- z galaxies. *Mon. Not. R. Astron. Soc.* **362**, 799–825 (2005).
- Cardelli, J. A., Clayton, G. C. & Mathis, J. S. The relationship between infrared, optical, and ultraviolet extinction. *Astrophys. J.* **345**, 245–256 (1989).
- Calzetti, D. *et al.* The dust content and opacity of actively star-forming galaxies. *Astrophys. J.* **533**, 682–695 (2000).
- Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee: the MCMC hammer. *Publ. Astron. Soc. Pacif.* **125**, 306–312 (2013).
- Kriek, M. *et al.* An ultra-deep near-infrared spectrum of a compact quiescent galaxy at $z = 2.2$. *Astrophys. J.* **700**, 221–231 (2009).

Greenland was nearly ice-free for extended periods during the Pleistocene

Joerg M. Schaefer^{1,2}, Robert C. Finkel^{1,3}, Greg Balco⁴, Richard B. Alley⁵, Marc W. Caffee⁶, Jason P. Briner⁷, Nicolas E. Young¹, Anthony J. Gow⁸ & Roseanne Schwartz¹

The Greenland Ice Sheet (GIS) contains the equivalent of 7.4 metres of global sea-level rise¹. Its stability in our warming climate is therefore a pressing concern. However, the sparse proxy evidence of the palaeo-stability of the GIS means that its history is controversial (compare refs 2 and 3 to ref. 4). Here we show that Greenland was deglaciated for extended periods during the Pleistocene epoch (from 2.6 million years ago to 11,700 years ago), based on new measurements of cosmic-ray-produced beryllium and aluminium isotopes (¹⁰Be and ²⁶Al) in a bedrock core from beneath an ice core near the GIS summit. Models indicate that when this bedrock site is ice-free, any remaining ice is concentrated in the eastern Greenland highlands and the GIS is reduced to less than ten per cent of its current volume. Our results narrow the spectrum of possible GIS histories: the longest period of stability of the present ice sheet that is consistent with the measurements is 1.1 million years, assuming that this was preceded by more than 280,000 years of ice-free conditions. Other scenarios, in which Greenland was ice-free during any or all Pleistocene interglacials, may be more realistic. Our observations are incompatible with most existing model simulations that present a continuously existing Pleistocene GIS. Future simulations of the GIS should take into account that Greenland was nearly ice-free for extended periods under Pleistocene climate forcing.

The possibility that future warming will cause destabilization of the GIS has motivated the use of geological records to estimate the climate sensitivity of the GIS. Terrestrial studies^{2,3} have argued that the palaeo-environment of the Kap København Formation in north Greenland implied an ice-free Greenland, with temperatures nearly 6 °C above present persisting for about 20,000 years (20 kyr) from 1.8 million years (Myr) ago to 2.0 Myr ago. Marine sedimentary proxy data from sites off southwest Greenland^{5,6} are interpreted to indicate a smaller GIS during both the Marine Isotope Stage (MIS) 5e (or Eemian; about 120 kyr ago) and MIS 11 (about 410 kyr ago) interglacial periods. Biomolecules in basal ice of the Dye-3 ice core in southern Greenland provide evidence for subarctic conditions (and thus a smaller GIS) sometime in the past million years or so⁷ and a recent review⁸ argues that the near-field and far-field data require major ice-sheet fluctuations, and allow (but do not require) near-total ice loss during the Pleistocene. On the other hand, data from the basal NEEM ice core⁹ indicate minor ice-surface lowering during MIS 5e despite temperatures several degrees warmer than present. The geochemistry of Greenland Ice Sheet Project Two (GISP2) silty basal ice has been interpreted as being consistent with the scenario of continuous ice cover for the past 2.6 Myr (ref. 4) and trapped air enclosed in the silty ice layer of the nearby Greenland Ice Sheet Project (GISP) core indicate basal ice ages exceeding 1 Myr (ref. 10; Methods).

The GIS survived mid-Holocene temperatures somewhat warmer than those of the past millennium and many model simulations show a relatively stable GIS over the interglacials of the recent geologic past^{11,12}. However, simulations also show that the warming required to remove most of the GIS is model-dependent and sensitive to external forcings and internal feedbacks, including insolation forcing, accumulation rate parameterization, and distribution and seasonality of temperature. Results imply temperature thresholds for ice-sheet stability between one¹¹ and a few degrees Celsius above present temperatures (see review in ref. 8; also refs 13 and 14). Because the GIS sensitivity probably changes with increasing forcing temperature, model time-scales for ice-sheet removal depend on the amplitude of the forcing: a temperature threshold of 2 °C with a 5,000-year response time given 3 °C warming was inferred by ref. 13, but more extreme temperature forcing allows for GIS removal within a few thousand or even several hundred years¹³. Thus, current model results remain ambiguous but do show that both the magnitude and the duration of warmth are important to ice-sheet deglaciation.

Overall, existing geological data and model experiments have not resolved the question of whether the GIS disappeared or shrank substantially in warm interglacial periods. Much of this uncertainty reflects the fact that the geological data mostly comprise inference from remote proxy records, since direct evidence, if it exists, is buried beneath the present ice sheet. Here we attempt to overcome this obstacle via cosmogenic nuclide analysis of sub-GIS bedrock.

On 1 July 1993, after five years of drilling and recovery of a 3,040.3-m-long ice core and a 13.1-m-long core of sediment-rich basal ice, the GISP2 project penetrated 1.55 m of bedrock¹⁵ (Figs 1 and 2). We describe measurements of cosmic-ray-produced in-situ ¹⁰Be and ²⁶Al from this GISP2 bedrock core. ¹⁰Be and ²⁶Al, with half-lives of 1.4 Myr (refs 16 and 17) and 0.7 Myr (ref. 18) respectively, are trace radionuclides produced in situ by nuclear interactions between cosmic-ray particles and rocks exposed at Earth's surface. The cosmic-ray flux decreases exponentially with an e-folding length ($1/e \approx 0.37$) of about 60 cm in rock or about 1.5 m in ice, so cosmogenic-nuclide production is negligible beneath ice sheets. The presence of any substantial *in situ* cosmogenic radionuclide concentration in subglacial bedrock indicates geologically recent near-surface exposure and thus ice-free conditions. Pioneering analysis in the 1990s, published as an abstract¹⁹, indicated detectable ¹⁰Be and ²⁶Al in the GISP2 bedrock core, but overall uncertainties remained large enough to prevent unambiguous conclusions about past GIS change. Here we describe comprehensive new ¹⁰Be and ²⁶Al measurements, a detailed analysis of the data, and their implications for past GIS dynamics.

After separating and purifying quartz from segments of bedrock core (Methods), we combined aliquots into the largest number of depth

¹Lamont-Doherty Earth Observatory, Geochemistry, Palisades, New York 10964, USA. ²Department of Earth and Environmental Sciences, Columbia University, New York, New York 10027, USA.

³Department of Earth and Planetary Sciences, University of California, Berkeley, Berkeley, California 94720, USA. ⁴Berkeley Geochronology Center, 2455 Ridge Road, Berkeley, California

94709, USA. ⁵Department of Geosciences, Pennsylvania State University, University Park, PA 16802, USA. ⁶Department of Physics and Astronomy, Purdue University, 525 Northwestern Avenue, West Lafayette, Indiana 47907, USA. ⁷Department of Geology, University at Buffalo, Buffalo, New York 14260, USA. ⁸US Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire 03755, USA.

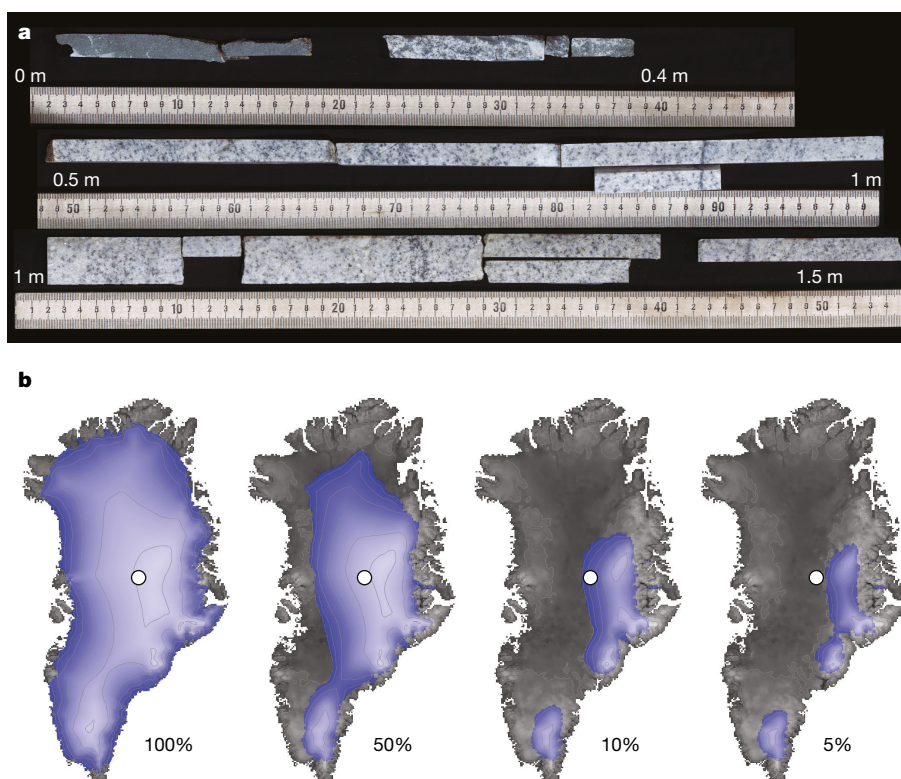


Figure 1 | The GISP2 bedrock core and GIS deglaciation. **a**, The GISP2 bedrock core. Note the lithologic transition from the top piece (carbonate) to granitoid and the excellent condition of the core (photo copyright Geoffrey Hargreaves, National Ice Core Laboratory, Denver, 2015). **b**, Snapshots of GIS deglaciation ranging from 100% to 5% ice volume relative to today, from simulations by the Penn State University Ice Sheet

Model (PSU) coupled to the Victoria Earth System Model (UVic ESCM). The white circle indicates the GISP2 site. Similar model simulations, such as given in ref. 26, reveal this as a robust spatial deglaciation pattern of the GIS and show that when the GISP2 site is ice-free, only small ice remnants in the east Greenland highlands exist. Image adapted from ref. 27, Springer-Verlag.

increments consistent with maintaining the minimum sample size needed for accurate measurements (Fig. 1). This yielded in situ ^{10}Be measurements for five distinct segments (Fig. 2). For ^{26}Al , we further combined samples to make two amalgamated samples for precise $^{26}\text{Al}/^{10}\text{Be}$ measurements (GISP2/3 combined with GISP4 and GISP5/6 combined with GISP7; see Fig. 2, Methods and Extended Data Tables 1–3).

^{10}Be concentrations range from $9,800 \pm 490$ atoms g^{-1} to $24,800 \pm 740$ atoms g^{-1} and ^{26}Al concentrations are $54,900 \pm 5,400$ atoms g^{-1} to $88,000 \pm 5,200$ atoms g^{-1} (Fig. 2; Extended Data Tables 2 and 3; errors reported here and elsewhere are 1σ). These are one to two orders of magnitude above blank levels and three to four orders of magnitude above the concentration expected from cosmogenic-nuclide production by deeply penetrating cosmic-ray muons at the base of a 3-km-thick ice sheet^{20,21} (Methods).

The measured in situ $^{26}\text{Al}/^{10}\text{Be}$ ratios are 4.2 ± 0.3 in the top sample and 4.1 ± 0.6 in the lower sample (Extended Data Table 1), significantly below the surface production ratio²² of 6.75. This indicates that the period of surface exposure recorded by ^{10}Be and ^{26}Al concentrations was followed by a considerable period of time during which samples were deeply enough buried, presumably by the GIS, to stop the cosmic-ray flux (for example, see ref. 23) and let the $^{26}\text{Al}/^{10}\text{Be}$ ratio ‘decay’ during burial, with an apparent half-life of 1.4 Myr (Methods). These $^{26}\text{Al}/^{10}\text{Be}$ ratios limit the duration of continuous ice cover of the GISP2 bedrock to a maximum duration of 1.1 ± 0.1 Myr.

The measured in situ ^{10}Be and ^{26}Al concentrations themselves and their variation with depth mean that the GISP2 site must have been ice-free for a considerable period of time in the recent geologic past. The approximate e-folding length of the measured ^{10}Be concentration depth profile is nearly 260 g cm^{-2} (Methods), which requires contributions both from high-energy neutron spallation (predominant at the surface, decreasing with an e-folding length near 140 g cm^{-2}) and muon

interactions (dominant below a depth of a few metres, with an e-folding length exceeding $1,500 \text{ g cm}^{-1}$) (refs 2 and 24). This both shows that samples were close to the surface when exposed to the cosmic-ray flux, and allows us to estimate their depth below the surface at that time (Methods). The data are best explained by $350 \pm 20 \text{ g cm}^{-2}$ of cover above the present bedrock surface, that is, about 1.3 m of rock, about 2 m of soil, or about 4 m of ice (Fig. 2; Methods). Thus, not even a thin ice cap could have been present at this site during the period of exposure. In addition, our data limit possible subglacial erosion at the GISP2 site to no more than 1–2 m during the period the site has been covered by the ice sheet, consistent with long-term⁴ and modern observations²⁵ (Methods). Given this thickness of cover of the bedrock surface during exposure, the shortest exposure period consistent with the measurements is 280 ± 30 kyr (Methods).

Simulations of Greenland deglaciation in warm climates consistently predict the GISP2 site to be one of the last parts of the ice sheet to disappear; when it is ice-free, only a small ice cap in the eastern highlands remains^{26,27} (Fig. 1; Methods), preserving old ice¹⁰. This implies that when GISP2 bedrock was exposed to the surface cosmic-ray flux, more than 90% of the entire GIS was absent.

Although we cannot identify a unique ice-cover history, the ^{10}Be and ^{26}Al data can be used to test specific Pleistocene GIS change scenarios that are or are not consistent with the data (Fig. 3). The simplest scenario consistent with the data, which also provides an upper limit on the length of time the present GIS has been stable, consists of more than 280 ± 30 kyr of ice-free conditions followed by 1.1 ± 0.1 Myr of continuous ice cover (the maximum-stability scenario ‘1’ in Fig. 3; see Methods). If the ice sheet had been continuously present for longer, the $^{26}\text{Al}/^{10}\text{Be}$ ratios would be lower than observed (for example, continuous ice cover over the past 2.6 Myr implies an $^{26}\text{Al}/^{10}\text{Be}$ ratio no higher than 2).

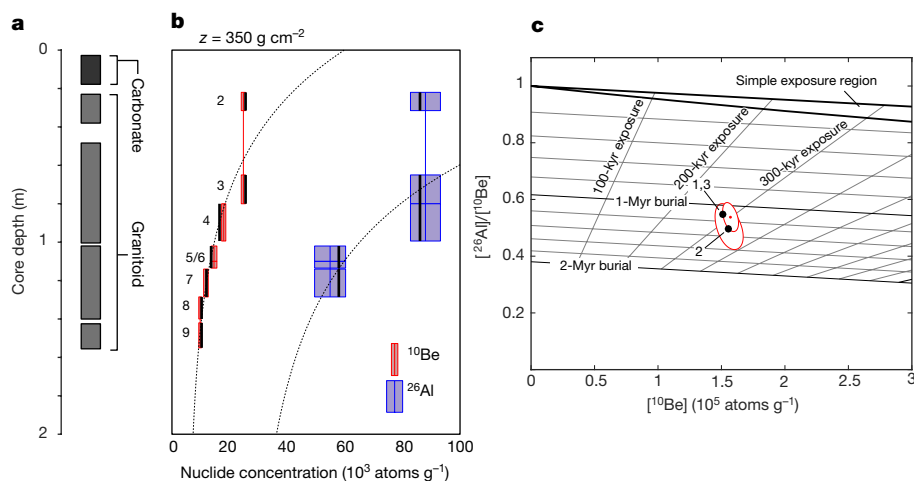


Figure 2 | ^{10}Be and ^{26}Al data for the GISP2 bedrock. **a**, Schematic core stratigraphy showing core recovery. Image adapted from ref. 15, International Glaciological Society. **b**, ^{10}Be and ^{26}Al concentrations in core segments, numbered in black. Heights of boxes correspond to the length of core segments. Thin vertical lines connect stacked samples that we amalgamated for single measurement where necessary. Specifically, for ^{10}Be measurements, we combined the individual core segments as follows: (1) GISP2 and GISP3; (2) GISP4; (3) GISP5 and GISP6; (4) GISP7; (5) GISP8 and GISP9; for ^{26}Al we combined GISP2, GISP3 and GISP4 for the analysis of the top sample and GISP5, GISP6 and GISP7 for the lower sample; please note that GISP5 and GISP6, and GISP8 and GISP9 are difficult to detect in the figure as individual samples. Widths of boxes show 1σ measurement uncertainties. Black dashed lines show predicted nuclide concentrations as a function of depth for the best-fitting

However, this maximum-stability scenario does not resemble available palaeoclimate data²⁸ (Fig. 3; Methods). Other, more complex GIS change scenarios are equally consistent with our observations, subject to the constraints that the most recent ice-free period ended no longer than 1.1 ± 0.1 Myr ago, and that the smaller the

bedrock cover thickness (350 g cm^{-2}) and the maximum-stability GIS exposure–burial scenario (‘1’ in Fig. 3). Heavy black lines are predicted nuclide concentrations averaged over depth ranges corresponding to each measurement; our fitting procedure (see Methods) minimizes the mismatch between these and the measurements. **c**, Two-nuclide diagram comparing measured ^{10}Be and ^{26}Al concentrations, normalized to production rates beneath 350 g cm^{-2} cover of the bedrock surface, with predicted concentrations for various exposure histories (see ref. 29 for a complete explanation of this kind of diagram). Red ellipses are 68% confidence intervals on the measurements, considering measurement uncertainty only. Grid shows isolines of exposure and burial time for idealized two-stage exposure histories. Black dots show predicted nuclide concentrations for exposure histories, numbered as in Fig. 3.

number of distinct ice-free periods, the longer the duration of each period must be. For example, an opposite endmember that implies minimal long-term GIS stability is an episodic steady-state scenario in which Greenland was ice-free for 8 kyr of each 100-kyr glacial–interglacial cycle spanning several million years (the minimum-stability scenario ‘2’ in Fig. 3). More realistic intermediate scenarios in which Greenland was ice-free for several thousand years during numerous (but not all) major Pleistocene interglacials, including MIS 11 as hypothesized from marine proxy records^{5,6}, are also consistent with the measurements (the multiple-exposures scenario ‘3’ in Fig. 3).

Our observations preclude any scenario in which the GIS was continually present throughout the entire Pleistocene. The data require ice-free periods in Greenland during the interval of 41-kyr-long glacial–interglacial cycles and permit (but do not require) subsequent ice-free periods during the Mid-Pleistocene Transition (MPT) or the interval of 100-kyr-long glacial–interglacial cycles (Fig. 3).

To summarize, direct and robust evidence from the GISP2 bedrock core shows that the GIS was almost completely absent for an extended period of time during the Pleistocene. Our results do not directly determine the ice-dynamical processes responsible for the GIS deglaciation, but this first-order result is incompatible with many existing ice-sheet models or their respective climate-driving scenarios and provides important constraints for future simulations of past and future changes of the GIS. Models driven by boundary conditions appropriate to the warmest and most pronounced Pleistocene interglaciations must simulate the near-total disappearance of the ice sheet.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 June; accepted 5 October 2016.

1. Houghton, J. T. et al. (eds) *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)* (Cambridge Univ. Press, 2001).

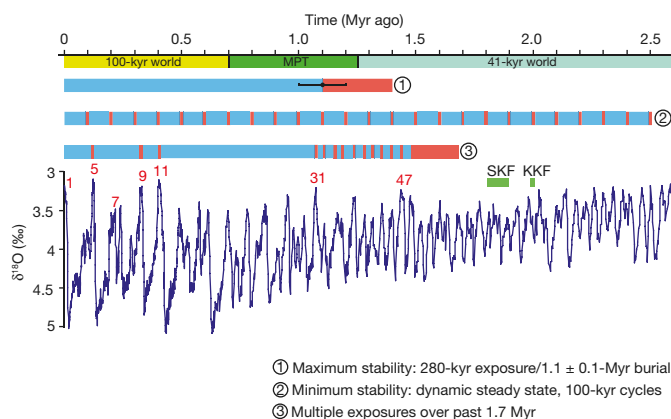


Figure 3 | Exposure–burial scenarios consistent with the data.

Exemplary exposure–burial scenarios for the GISP2 bedrock core over the past 2.6 Myr that are consistent with the data, plotted over the climate record derived from stacked benthic foraminifera³⁰, a proxy for global ice volume and deep-ocean temperature. The horizontal bars show three such scenarios consistent with observations for the GISP2 bedrock site, including the maximum-stability GIS scenario ‘1’, with red periods representing exposure of the GISP2 location and blue periods representing burial by the GIS overlying the GISP2 site. Small numbers in red above the climate curve indicate prominent interglacial marine isotope stages. The MPT (0.7–1.2 Myr ago) is the transition period from the early Pleistocene (41-kyr glacial–interglacial periodicity) to the mid-/late Pleistocene (100-kyr glacial–interglacial periodicity). The Store Koldewey Formation² (SKF) and the Kap København Formation^{2,3} (KKF) are Quaternary sediment formations in northern Greenland, indicating that Greenland was nearly ice-free at that time.

2. Bennike, O. *et al.* Early Pleistocene sediments on Store Koldewey, northeast Greenland. *Boreas* **39**, 603–619 (2010).
3. Funder, S., Bennike, O., Bocher, J., Israelson, C. & Petersen, K. S. Late Pliocene Greenland—the Kap Kobenhavn Formation in North Greenland. *Bull. Geol. Soc. Den.* **48**, 117–134 (2001).
4. Bierman, P. R. *et al.* Preservation of a preglacial landscape under the center of the Greenland Ice Sheet. *Science* **344**, 402–405 (2014).
5. de Vernal, A. & Hillaire-Marcel, C. Natural variability of Greenland climate, vegetation, and ice volume during the past million years. *Science* **320**, 1622–1625 (2008).
6. Reyes, A. V. *et al.* South Greenland ice-sheet collapse during Marine Isotope Stage 11. *Nature* **510**, 525–528 (2014).
7. Willerslev, E. *et al.* Ancient biomolecules from deep ice cores reveal a forested Southern Greenland. *Science* **317**, 111–114 (2007).
8. Alley, R. B. *et al.* History of the Greenland Ice Sheet: paleoclimatic insights. *Quat. Sci. Rev.* **29**, 1728–1756 (2010).
9. NEEM Community Members. Eemian interglacial reconstructed from a Greenland folded ice core. *Nature* **493**, 489–494 (2013).
10. Yau, A. M., Bender, M. L., Blunier, T. & Jouzel, J. Setting a chronology for the basal ice at Dye-3 and GRIP: implications for the long-term stability of the Greenland Ice Sheet. *Earth Planet. Sci. Lett.* **451**, 1–9 (2016).
11. Calov, R., Robinson, A., Perrette, M. & Ganopolski, A. Simulating the Greenland ice sheet under present-day and palaeo constraints including a new discharge parameterization. *Cryosphere* **9**, 179–196 (2015).
12. Otto-Bliesner, B. L. & Brady, E. C. The sensitivity of the climate response to the magnitude and location of freshwater forcing: last glacial maximum experiments. *Quat. Sci. Rev.* **29**, 56–73 (2010).
13. Applegate, P. J., Parizek, B. R., Nicholas, R. E., Alley, R. B. & Keller, K. Increasing temperature forcing reduces the Greenland Ice Sheet's response time scale. *Clim. Dyn.* **45**, 2001–2011 (2015).
14. Stone, E. J., Lunt, D. J., Rutt, I. C. & Hanna, E. Investigating the sensitivity of numerical model simulations of the modern state of the Greenland ice-sheet and its future response to climate change. *Cryosphere* **4**, 397–417 (2010).
15. Gow, A. J. & Meese, D. A. Nature of basal debris in the GISP2 and Byrd ice cores and its relevance to bed processes. In *Proceedings of the International Symposium on Glacial Erosion and Sedimentation* (ed. Collins, D.) 134–140 (International Glaciological Society, 1996).
16. Chmeleff, J., von Blanckenburg, F., Kossert, K. & Jakob, D. Determination of the Be-10 half-life by multicollector ICP-MS and liquid scintillation counting. *Nucl. Instrum. Methods Phys. Res. B* **268**, 192–199 (2010).
17. Korschinek, G. *et al.* A new value for the half-life of Be-10 by heavy-ion elastic recoil detection and liquid scintillation counting. *Nucl. Instrum. Methods Phys. Res. B* **268**, 187–191 (2010).
18. Nishiizumi, K. *et al.* Absolute calibration of ¹⁰Be AMS standards. *Nucl. Instrum. Methods Phys. Res. B* **258**, 403–413 (2007).
19. Nishiizumi, K. *et al.* In situ produced cosmogenic nuclides in GISP2 rock core from Greenland Summit. *Eos* **77**, F428, abstr. OS41B-10 (1996).
20. Heisinger, B. *et al.* Production of selected cosmogenic radionuclides by muons: 2. Capture of negative muons. *Earth Planet. Sci. Lett.* **200**, 357–369 (2002).
21. Heisinger, B. *et al.* Production of selected cosmogenic radionuclides by muons. 1. Fast muons. *Earth Planet. Sci. Lett.* **200**, 345–355 (2002).
22. Balco, G. & Rovey, C. W. An isochron method for cosmogenic-nuclide dating of buried soils and sediments. *Am. J. Sci.* **308**, 1083–1114 (2008).
23. Granger, D. E. & Muzikar, P. F. Dating sediment burial with in situ-produced cosmogenic nuclides: theory, techniques, and limitations. *Earth Planet. Sci. Lett.* **188**, 269–281 (2001).
24. Marrero, S. M. *et al.* Cosmogenic nuclide systematics and the CRONUScal program. *Quat. Geochronol.* **31**, 160–187 (2016).
25. Bender, M. L., Burgess, E., Alley, R. B., Barnett, B. & Clow, G. D. On the nature of the dirty ice at the bottom of the GISP2 ice core. *Earth Planet. Sci. Lett.* **299**, 466–473 (2010).
26. Alley, R. B., Clark, P. U., Huybrechts, P. & Joughin, I. Ice-sheet and sea-level changes. *Science* **310**, 456–460 (2005).
27. Fyke, J., Eby, M., Mackintosh, A. & Weaver, A. Impact of climate sensitivity and polar amplification on projections of Greenland Ice Sheet loss. *Clim. Dyn.* **43**, 2249–2260 (2014).
28. Dutton, A. *et al.* Sea-level rise due to polar ice-sheet mass loss during past warm periods. *Science* **349**, (2015).
29. Granger, D. E. in *In-Situ Produced Cosmogenic Nuclides And Quantification Of Geological Processes* (eds Siame, L., Bourles, D. & Brown, E. T.) Spec. Pap. 415, 1–16 (The Geological Society of America, 2006).
30. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic delta O-18 records. *Paleoceanography* **20**, PA1003 (2005).

Acknowledgements We gratefully acknowledge help from the National Ice Core Laboratory (NICL) and thank the GISP2 steering committee for providing the bedrock core samples. J.M.S. acknowledges support by the Lamont Climate Center and the Comer Family Foundation. R.B.A. acknowledges support by the NSF (AGS 1338832), as do J.M.S., J.P.B. and N.E.Y. (PLR Arctic System Science Program number 1503959). We thank J. Fyke for assistance in producing Fig. 1b. G.B. acknowledges support from the Ann and Gordon Getty Foundation. M.W.C. acknowledges support from the US National Science Foundation, grant EAR-1153689. The pioneering cosmogenic-nuclide study of the GISP2 bedrock core under the lead of Kuni Nishiizumi in the late 1990s motivated our study. This is LDEO publication number 8068.

Author Contributions J.M.S. and R.C.F. initiated and coordinated the project, were in charge of the data production and wrote the first draft of the manuscript. G.B. provided the data analysis. R.B.A. participated in the GISP2 project, and provided glaciological expertise and model perspective. A.J.G. was part of the first scientist team at the GISP2 camp when the bedrock core was retrieved, examined the rock core and provided stratigraphic ice-bedrock expertise. M.W.C. measured the samples for Al isotopes. R.S. processed all the rock samples and did the Al and Be extraction. J.P.B. and N.E.Y. provided Arctic glacier expertise and prepared final figures. All authors read and edited multiple versions of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.M.S. (schaefer@ldeo.columbia.edu).

Reviewer Information Nature thanks P.-H. Blard, G. Milne and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Analytical methods. We received either half- or quarter-splits of depth intervals from the 3.4-cm-diameter core (Fig. 1). After separating a quartz fraction from each core segment, in most cases we combined quartz fractions from multiple core segments to yield sufficient quartz for ^{10}Be and ^{26}Al analyses. Quartz separation, purification and Be and Al extraction from quartz then followed standard procedures used at the Lamont Cosmogenic Nuclide Laboratory³¹. After Be and Al extraction, we further combined Al fractions to obtain sufficient Al for ^{26}Al measurements. Thus, more core segments were combined for ^{26}Al analysis than for ^{10}Be , resulting in fewer ^{26}Al analyses overall (Fig. 2). Note that we are not including in further discussion a preliminary ^{26}Al measurement from the lowest core segments (GISP8 and GISP9), because we could not verify the Al yield in the extraction chemistry (Extended Data Table 3). In view of the importance of these specific samples and the possibility that the concentrations would be close to background levels, we added additional quality-control procedures. These include expanded monitoring of process blanks (Extended Data Tables 2 and 3), time-progressive blank corrections for the aluminium isotope measurements (see below) and ICP-OES standardization by in-house, gravimetrically prepared ^{27}Al standards cross-checked with commercial standards. Be and Al isotope ratio measurements by accelerator mass spectrometry (AMS) were carried out at the Center for Accelerator Mass Spectrometry at the Lawrence Livermore National Laboratory³² and PRIME Laboratory, Purdue University, respectively.

$^{26}\text{Al}/^{27}\text{Al}$ analysis at PRIME Laboratory. Until recently, most AMS measurements of the $^{26}\text{Al}/^{27}\text{Al}$ ratio used Al^- as the secondary ion. Injecting Al^- has the advantage that Mg^- is not formed, so ^{26}Mg , the main interference for ^{26}Al measurements, is not present. However, Al^- currents are low (typically $1\ \mu\text{A}$ or less). Injecting AlO^- improves beam current by a factor of ten³³, thereby increasing the number of ^{26}Al counts measured at the detector. However, MgO^- is also produced, injected and accelerated, allowing ^{26}Mg to reach the detector. The gas-filled magnet installed at PRIME Laboratory allows effective suppression of this ^{26}Mg interference³⁴. Specifically, N_2 at low pressure is bled into the beam line as it passes between the poles of a specially designed magnet placed before the traditional dE/dx detector³⁵. ^{26}Mg and ^{26}Al ions have different charge-exchange cross-sections with N_2 . Thus they have different charge states as they traverse the curved part of the flight tube passing through the magnet and therefore have different radii of curvature. An aperture at the magnet exit then suppresses ^{26}Mg ions by a factor of about 10^4 while allowing ^{26}Al into the detector. The ^{26}Al and some residual ^{26}Mg enter the detector. For low count rates ($<1,000$ counts per second), the dE/dx detector is able to distinguish these ions unambiguously on the basis of their energy-loss properties. The gas-filled magnet allows much greater secondary ion currents and allows the precision of $^{26}\text{Al}/^{27}\text{Al}$ measurements to approach that of $^{10}\text{Be}/^9\text{Be}$ measurements³⁶.

For $^{26}\text{Al}/^{27}\text{Al}$ measurement the tandem is run at a terminal voltage of 8.5 MV. Al ions in charge state $+7$ with an energy of around 65 MeV are selected. The ^{26}Al loses about a third of its energy traversing the gas-filled magnet, the remaining two-thirds being available for particle identification in the dE/dx detector. The samples analysed for this work had uniformly low Mg count rates, so the ^{26}Al particle peak was well resolved from other peaks in the energy spectra. Maximum Al-beam currents were near or above $10\ \mu\text{A}$ with currents remaining above $5\ \mu\text{A}$ even after an hour of sputtering. All the samples were mixed with niobium before loading into sample cathodes, but the combined sample GISP2/3/4 was measured after PRIME Laboratory switched to niobium as the binder for the AMS standards, which proved to be superior and cleaner in the AMS than the AMS standards used in the earlier measurements of the GISP2 samples that were mixed with silver. This might be one explanation for the considerably lower blank-level corresponding to the GISP2/3/4 measurement (Extended Data Table 3).

$^{26}\text{Al}/^{27}\text{Al}$ background correction for the GISP2 samples. The corrections to the $^{26}\text{Al}/^{27}\text{Al}$ 'raw' ratios of our samples are less than 2% for the top sample GISP2/3/4 and 15% for the lower GISP5/6/7 sample (Extended Data Table 3). In addition to the traditional background correction by simply subtracting the mean ratio of the measured process blanks from the $^{26}\text{Al}/^{27}\text{Al}$ ratio of the samples, we chose to apply a time-progressive background correction that takes into account the different run times of the two samples and the blanks. This can be important, because our experience indicates that the most likely source of the ^{26}Al blank is ion source memory, rather than addition of ^{26}Al during chemical processing. The ^{26}Al machine memory derived from higher ratio samples, such as standards, previously measured in the ion source, is a function of measurement time, and thus samples measured for longer accumulate higher levels of memory ^{26}Al . We monitored the machine memory by measuring multiple blanks for periods of time comparable to the periods for which the corresponding samples were measured (Extended Data Table 3). The number of mass 26 counts in the blanks was converted into counts of ^{26}Al per minute of run time and this count rate was subtracted from

the raw count rate of the samples measured at about the same time. The total error in the ratio based on this correction reflects the uncertainties of the sample measurement together with the uncertainties of this correction. We note that the two correction procedures applied here yield very similar results (Extended Data Table 3).

Data analysis. We describe how we infer information about the exposure history of GISP2 subglacial bedrock from the measured ^{26}Al and ^{10}Be concentrations. The three basic characteristics of these data are: (1) cosmogenic-nuclide concentrations decrease with depth in the core; (2) concentrations decrease less rapidly than we would expect if the bedrock surface in the core had been the land surface during exposure; and (3) the measured $^{26}\text{Al}/^{10}\text{Be}$ ratio in these samples is substantially less than the surface production ratio of 6.75. The change in nuclide concentration with depth in the core (as described in detail below) is diagnostic of the depth below the surface at which production took place. The $^{26}\text{Al}/^{10}\text{Be}$ ratio is important because these nuclides are produced at a fixed ratio during surface exposure ($^{26}\text{Al}/^{10}\text{Be} = 6.75$; ref. 22), but ^{26}Al has a shorter half-life (0.7 Myr; ref. 18) than ^{10}Be does (1.4 Myr; refs 16 and 17). A sample that has experienced a single period of surface exposure has a $^{26}\text{Al}/^{10}\text{Be}$ ratio that conforms to the production ratio, but if the sample is then buried sufficiently deeply to halt additional nuclide production, the $^{26}\text{Al}/^{10}\text{Be}$ ratio decreases over time with a half-life of 1.4 Myr.

It is not possible to invert our observations to yield a single unique exposure history because there exist an infinite number of exposure histories that are consistent with these observed characteristics. Thus, we here propose a number of broad classes of exposure histories and ask if they can be fitted to the observations. For example, if we propose that the observations can be explained by a single period of surface exposure followed by a single period of ice cover, we can ask whether there exist any values for the exposure and burial durations that provide an acceptable fit to the observations. If no such values exist, this class of exposure histories is not consistent with the observations. If such values do exist, they define exposure histories that are consistent with the observations and thus could potentially represent the true exposure history of the site.

We will use this reasoning to support three main conclusions. First, although the present bedrock surface in the core must have been covered by some additional rock, soil or ice during past periods of cosmic-ray exposure, the thickness of this material cannot have been more than a few metres. Thus, the observations require the absence of the GIS during periods of cosmic-ray exposure. Second, the longest possible length of time that the present ice sheet could have remained continuously present at the core site is approximately 1.1 Myr. Third, although there exist many scenarios of ice-sheet change that are consistent with the observations, all these scenarios require the absence of the ice sheet at the core site during some or all Pleistocene interglaciations.

We now discuss the depth-dependence of the cosmogenic-nuclide concentrations. Production of ^{10}Be and ^{26}Al occurs both by high-energy neutron spallation and by muon interactions. At and immediately below the surface, most production is by spallation, and spallogenic production rates decrease exponentially with depth with an e-folding length near $140\ \text{g cm}^{-2}$ (this value is appropriate for high latitude). With increasing depth, spallogenic production becomes less important than production due to deeply penetrating muons, which has a longer e-folding length ($>1,500\ \text{g cm}^{-2}$, increasing with depth²⁴). A single exponential curve fit to our observed nuclide concentrations has an e-folding length intermediate between these two ($262\ \text{g cm}^{-2}$, although because a single exponential function does not fit the data adequately, this value is not physically meaningful). Thus, the observed nuclide concentrations reflect both spallogenic and muon production. In other words, the observations are not consistent with a single period of exposure during which the present bedrock surface was the land surface (in this case the measured concentrations would decrease with an e-folding length near $140\ \text{g cm}^{-2}$). The fact that we observe an apparent e-folding length that is longer than expected for spallogenic production alone requires that a large fraction of the measured nuclide concentrations is the result of production by muons, which means that production took place farther below the surface than the samples are now below the bedrock surface. The data are also not consistent with exposure more than a few metres below the surface (in which case the contribution from spallogenic production would be negligible, and the concentrations would decrease with an e-folding length characteristic of muon production).

Extended Data Fig. 1 shows that the depth dependence of the production rate immediately below the surface has a much shorter e-folding length than has the depth dependence of the measured nuclide concentrations. In other words, an exposure model including a period of surface exposure, with negligible erosion, during which time the bedrock surface was the land surface, does not fit the data. We also show the result of assuming that the bedrock surface was covered by additional mass during the period of exposure. In other words, we are trying to fit the measurements to the production profile by adjusting how far the bedrock

surface was below the actual land surface at the time that exposure took place. We represent this by an additional mass cover thickness covering the bedrock surface. The best-fitting cover thickness, the fit of which is shown in Fig. 2 and also in Extended Data Fig. 1, is 350 g cm^{-2} . A nominal (Monte Carlo) uncertainty on this value derived only from the measurement uncertainties is $\pm 20 \text{ g cm}^{-2}$. However, the best-fitting cover thickness also depends on the assumed mean atmospheric pressure at the site during exposure (which, as discussed in more detail below, is not well constrained), because this affects the relative proportion of spallogenic and muon-induced production. For example, reducing the atmospheric pressure by 25 hPa (equivalent to an increase in site elevation of about 200 m) would imply a best-fitting cover thickness of 375 g cm^{-2} instead of 350 g cm^{-2} . By this argument, the nominal measurement uncertainty of 20 g cm^{-2} does not provide a physically meaningful error estimate.

350 g cm^{-2} of mass thickness is approximately 1.3 m of rock, 1.7 m of soil or 3.8 m of ice. Thus, the simplest explanation for the observed depth-dependence of the measured nuclide concentrations is that the bedrock surface was covered by 1–4 m of soil, sediment, rock or ice when exposure took place. Later, for completeness, we consider a more complicated and geologically less likely potential explanation involving a long period of continuous surface exposure at a relatively high erosion rate, and show that it does not provide as good a fit to the measurements. To summarize our discussion of the depth-dependence of the measured nuclide concentrations, (1) this shows that the present bedrock surface was probably covered by some additional shielding at the time of nuclide production, and (2) this requires that production took place within a few metres of Earth's surface. The observed nuclide inventory could not be the result of long residence under tens or hundreds of metres of ice; if it were, nuclide concentrations would decrease less with depth.

We now proceed by developing a forward model to compute the predicted nuclide concentrations in our samples given a specified exposure history. First, we describe our production-rate calculations in detail. We assume a reference ^{10}Be production rate due to spallation of $4.1 \text{ atoms g}^{-1} \text{ yr}^{-1}$, which follows ref. 37 but is adjusted for the difference in muon interaction cross-sections between those used here and those used in that work; a $^{26}\text{Al}/^{10}\text{Be}$ production ratio of 6.75 (ref. 22); and the production rate scaling method of ref. 38, as implemented in ref. 37. We assume that the land surface at the core site during periods of exposure is located at sea level, and compute the atmospheric pressure using the spatially variable atmosphere of ref. 37. We note that this is speculative because both the elevation of the site and the atmospheric pressure distribution would most probably be very different in the absence of the GIS; in addition, both would experience transient changes during ice-free periods. We take the ^{10}Be decay constant to be 4.99×10^{-7} (refs 16 and 17) and the ^{26}Al decay constant to be 9.83×10^{-7} (ref. 39). We compute production rates due to muons using the MATLAB implementation in ref. 37 of the method of refs 20 and 21. However, we use muon interaction cross-sections derived from fitting this method to measurements from a deep sandstone core in Antarctica. These measurements are reported in ref. 40; the fitted cross-sections are: for ^{10}Be , $f^* = 0.0011$ and $\sigma_0 = 0.81 \mu\text{b}$; for ^{26}Al , $f^* = 0.0084$ and $\sigma_0 = 13.6 \mu\text{b}$ (these symbols correspond to those used by refs 20 and 21). These differ from the values given in ref. 41, because we are using different code to compute the muon fluxes. Note that our main conclusions in this paper relate to the duration of ice cover of the site, which is sensitive to the values we use for the ^{10}Be and ^{26}Al decay constants and the $^{26}\text{Al}/^{10}\text{Be}$ production ratio, but is insensitive to inaccuracies in the absolute magnitude of the production rates. This is important because the uncertainty in our knowledge of the elevation and atmospheric pressure prevailing at the site during ice-free periods means that we are unlikely to be accurately estimating surface production rates when past exposure took place.

We now describe the model calculation of the nuclide concentrations expected for a particular exposure history. Assuming an exposure history is equivalent to specifying the mass cover thickness z_c (mass thickness has units of grams per square centimetre) above the surface of the bedrock in the core at all past time. We define a time coordinate t that is zero at the present time and positive for past times, and then define an exposure history as a function $z_c(t)$ that describes the variation in z_c over time. For example, if the exposure history consists of one period of surface exposure followed by one period of cover by the full thickness of the Greenland ice sheet, then $z_c(t)$ equals zero during the period of exposure and is effectively infinite (ice thickness of the order of 100 m or greater is effectively infinite burial, that is, zero cosmogenic-nuclide production) during the period of ice cover. Another example is that surface erosion during ice-free periods would be represented by $z_c(t)$ steadily decreasing during the ice-free period at the rate of erosion.

Given an exposure history $z_c(t)$ and also knowing the production rate of nuclide k as a function of mass depth z , $P_k(z)$ ($\text{atoms g}^{-1} \text{ yr}^{-1}$; calculated as described

above), the predicted average concentration $N_{j,k}$ (atoms g^{-1}) of nuclide k in sample j at the present time is given by the integral:

$$N_{j,k} = (z_{j,\text{bottom}} - z_{j,\text{top}})^{-1} \int_{z_{j,\text{top}}}^{z_{j,\text{bottom}}} \int_0^{t_{\text{max}}} P_k(\zeta + z_c(\tau)) \exp(-\lambda_k \tau) d\tau d\zeta \quad (1)$$

where $z_{j,\text{bottom}}$ and $z_{j,\text{top}}$ are the top and bottom mass depths (in units of g cm^{-2}) of sample j below the bedrock surface, $z_c(t)$ is the exposure history—that is, the mass thickness (in g cm^{-2}) covering the bedrock surface as a function of time t (in years)— λ_k is the decay constant (in yr^{-1}) of nuclide k , and τ and ζ are variables of integration. As noted above, the time coordinate is defined to be zero at the present time and positive for past times: t_{max} is the time that the exposure history begins, and the nuclide concentration at t_{max} is zero. In the case that the sample is composed of multiple core segments, we evaluate the integral for each segment separately and then use the weight of quartz contributed by each core segment to calculate the average nuclide concentration in the combined sample. We evaluate this integral using the default numerical integration routine in MATLAB software with, for computational efficiency, a piecewise-linear approximation of $P_k(z)$ defined on a logarithmic grid.

The misfit statistic we use in comparing predicted to measured nuclide concentrations is the chi-squared statistic χ^2 using the measurement uncertainty in the nuclide concentrations as the weighting. We note that because the measurement uncertainties for ^{10}Be are much smaller than those for ^{26}Al , this statistic assigns greater importance to the ^{10}Be measurements and is more sensitive to whether the model matches the depth profile of ^{10}Be concentrations than to whether the model matches $^{26}\text{Al}/^{10}\text{Be}$ ratios. A fitting statistic that weighted the ^{26}Al data set equally with the ^{10}Be data set might yield slightly different conclusions, although we tested this and did not find a large effect on the main conclusions of the paper. Another important point about the fit statistic is that we assume that the measured core depths and density are exact. This is unlikely to be strictly true, because the core is jointed and somewhat fragmented, with short missing sections (Fig. 1), and small vertical displacements of the core segments during recovery are likely. Thus, the probabilities of fit to the observations discussed below are likely to underestimate the model performance; if we accounted for additional uncertainties in depth and density we would obtain higher probabilities of fit.

We begin by fitting the simplest possible exposure history that can explain our data to the measurements. Several observations lead us to this simplest possible exposure history; by 'simplest', we mean the exposure history that has the fewest distinct periods of surface exposure or burial by the ice sheet. First, the fact that we observe substantial cosmogenic-nuclide concentrations that decrease with depth indicates that a period of surface exposure occurred. Second, the depth-dependence of the concentrations, as discussed above, indicates that the bedrock surface was probably shielded by additional cover at the time that exposure took place. Third, the fact that the observed $^{26}\text{Al}/^{10}\text{Be}$ ratio is lower than the surface production ratio indicates that the samples have been shielded from the surface cosmic-ray flux, presumably by the present GIS, for some time (Extended Data Table 1). These three observations show that the simplest possible exposure model that can explain our observations must include at least one period of surface exposure, at least one period of burial, and some thickness of additional cover of the bedrock surface. For the time being we assume zero surface erosion at all times. We refer to this model henceforth as 'model 1'.

Extended Data Fig. 2 shows the results of fitting model 1 with three adjustable parameters (the duration of a single period of exposure, the duration of a subsequent period of burial, and the cover thickness above the present bedrock surface) to the observations. Model 1 fits the observations with reduced χ^2 of 1.9 for 4 degrees of freedom, which is a probability of fit of 0.11. Given that the probability-of-fit is probably biased low as discussed above, we view this as an acceptable fit. As discussed above, note that the fitting statistics here are disproportionately affected by whether or not the depth profile fits the relatively high-precision ^{10}Be measurements; because the model is overparameterized in relation to a single $^{26}\text{Al}/^{10}\text{Be}$ pair, the exposure time and burial time parameters could be chosen to fit either pair exactly. The best-fitting exposure history here has 280 kyr of exposure under 350 g cm^{-2} of additional cover, followed by 1.1 Myr of burial. A Monte Carlo uncertainty analysis considering only measurement uncertainties indicates formal uncertainties in these parameters as follows: exposure time $282 \pm 31 \text{ kyr}$; burial time $1.110 \pm 0.092 \text{ Myr}$. If we include uncertainties in decay constants, the uncertainty on the burial time is 0.109 Myr . We note that these uncertainties only apply if model 1 is assumed to be correct a priori. In reality, of course, although model 1 is consistent with the cosmogenic-nuclide measurements taken by themselves, geological and palaeoclimate considerations show that it is unlikely. In addition, the best-fitting exposure duration

for this model is also somewhat dependent on the assumed atmospheric pressure. For example, decreasing the assumed atmospheric pressure by 25 hPa (as we did above in the discussion of cover thickness, which is approximately equivalent to a 200-m increase in elevation) decreases the best-fitting exposure duration by 6 kyr (to 276 kyr). However, this does not affect the estimate of burial duration.

Fitting model 1 highlights the fact that because of the additional cover thickness that we need to match the measured attenuation length of the ^{10}Be concentration depth profile, the exposure time to account for the observed nuclide concentrations must be on the order of 10^5 years. Taken out of this depth profile context, the measured nuclide concentrations are similar to typical surface concentrations in modern deglaciated landscapes that have experienced Holocene surface exposure only of the order of 10,000 years. However, because exposure took place below the land surface (below 1.3 m of rock, 1.7 m of soil or 3.4 m of ice; see above), they actually record an exposure history that is much longer than a single interglacial, or even a few interglacials.

For completeness, we now consider another theoretical explanation for the fact that the depth-dependence of the measured concentrations is inconsistent with spallogenic production alone. This could occur if the bedrock surface had been allowed to erode at a steady rate while continuously exposed at the surface for a long period of time. A long period of steady erosion means that a large fraction of the near-surface nuclide inventory was actually produced at depth by muons, and has been brought to the surface by erosion. This results in a longer apparent attenuation length near the surface than would be expected for spallogenic production only. In effect, steady erosion acts to drag the lower, longer-attenuation-length, part of the profile towards the surface. This scenario is unlikely for geological and palaeoclimatic reasons because it takes a long period of steady erosion for the muon-produced inventory to reach steady state. For example, at an erosion rate of 30 m Myr^{-1} (see below) it would take of the order of 1 Myr of continuous exposure and steady erosion for the muon-produced ^{10}Be inventory to reach steady state. In other words, although the steady-state scenario removes the need for additional sediment cover of the bedrock surface, it requires a very long period of undisturbed exposure before burial.

To evaluate this possibility, we fitted a model ('model 1B') with the following exposure history. First, the bedrock surface is the land surface—there is no sediment cover—and it erodes at a steady rate for long enough that nuclide concentrations reach a steady state in which production is balanced by nuclide loss via surface erosion and radioactive decay. Second, it is buried under ice for a period of time. Model 1B has two free parameters: the steady erosion rate and the burial duration. The best fit of model 1B to the data (Extended Data Fig. 2) has steady erosion at 32 m Myr^{-1} followed by 1.06 Myr of burial. This does not fit the data as well as model 1; the reduced χ^2 is 5.85 for 5 degrees of freedom, implying a probability-of-fit <0.0001 . Model 1B has difficulty matching both the depth dependence of the ^{10}Be concentrations and also the $^{26}\text{Al}/^{10}\text{Be}$ ratio. Perhaps more importantly, as discussed above, this scenario requires more than about 1 Myr of undisturbed steady erosion before burial, which is geologically unlikely in the light of evidence for glaciation of Greenland during the Pleistocene. In addition, the observation of high concentrations of meteoric ^{10}Be in sediment in GISP2 basal ice⁴ would be inconsistent with an erosion rate as high as 30 m Myr^{-1} . If the sediment analysed by ref. 4 has been buried under ice since 1.1 Myr ago as our $^{26}\text{Al}/^{10}\text{Be}$ ratios indicate, then a steady-state balance calculation based on Figure 3B of ref. 4 would suggest an erosion rate of 7 m Myr^{-1} , which would be typical of early Pleistocene cratonic landscapes in North America⁴². If the erosion rate were 7 m Myr^{-1} , we could not match the observed ^{10}Be depth profile with a steady-erosion model unless we also allowed for sediment cover of the bedrock surface. Thus, we conclude that model 1B is unrealistic and the most likely explanation for the observed depth-dependence of the ^{10}Be and ^{26}Al concentrations is that the present bedrock surface was covered by around 1–4 m of additional cover at the time that exposure took place.

To complete the discussion of two-stage exposure models, we observe that although we have considered two-stage models with no erosion and sediment cover (model 1) and steady-state erosion with no sediment cover (model 1B, which we have subsequently rejected), there exist intermediate scenarios that provide an acceptable fit to the data. Here we perform a Monte Carlo simulation in which we consider only two-stage exposure histories with a single period of exposure followed by a single period of burial, but allow the exposure duration, burial duration, erosion rate during exposure, and cover thickness to float. This model ('model 1C') is effectively overparameterized and finds many combinations of the parameters that fit the data acceptably well. Here we define 'acceptable' to be reduced $\chi^2 = 2.6$ for 3 degrees of freedom; that is, we cannot reject the model at 95% confidence. Extended Data Fig. 3 shows that a wide variety of combinations of exposure time, burial time and cover of the bedrock surface will adequately fit the observations. This exercise also demonstrates that the best-fitting burial duration is

only very weakly sensitive to the trade-off between exposure time, erosion rate and cover thickness. No matter what the exposure time, cover thickness and erosion rate before burial, the observed $^{26}\text{Al}/^{10}\text{Be}$ ratio constrains the burial time for any two-stage exposure history to be close to 1.1 Myr.

Finally, we now turn to exposure models with more than two stages. As has been discussed at length in the literature in the context of ^{26}Al – ^{10}Be concentrations observed in bedrock that has been covered by ice in the past (for example, see ref. 43), there are an infinite number of possible exposure-burial histories that will yield our observed ^{26}Al and ^{10}Be concentrations and ratio. In this case, however, all possible exposure histories share the constraint that at least one period of exposure must have ended no earlier than $1.1 (\pm 0.1)$ Myr ago. If this were not the case—if the present ice sheet at the core site had been in place for more than 1.1 Myr—the $^{26}\text{Al}/^{10}\text{Be}$ ratio in these samples would be lower than we observe at present. For example ref. 4 hypothesized that the GISP2 site has been continuously covered by the GIS since 2.7 Myr ago. If this hypothesis were correct (and if we also assume 350 g cm^{-2} of cover of the bedrock surface as discussed above, although this is not a critical assumption), the observed $^{26}\text{Al}/^{10}\text{Be}$ concentration in our samples could be no higher than 2.

In this section, therefore, we propose a series of representative multi-stage exposure-burial scenarios and determine whether or not they are consistent with the observations. We also use the observation from fitting model 1C (Extended Data Fig. 3) that the burial duration is not sensitive to the particular combination of erosion rate and cover of the bedrock surface during exposure periods, to limit the parameter space somewhat. Henceforth we will assume that the bedrock surface is covered by 350 g cm^{-2} of cover during exposure periods, and the surface erosion rate is zero. If this were not correct, that is, if the erosion rate were greater than zero during exposure periods, this assumption would lead us to underestimate the duration of exposure periods slightly. If we can take the meteoric ^{10}Be measurements of ref. 4 as a guide to the likely erosion rate as discussed above, this effect is not important to any of our conclusions. Note that the relevance of underestimating the erosion rate on the inferred duration of exposure periods can be estimated from Extended Data Fig. 3.

The first many-stage exposure history we consider is an endmember, opposite from the simplest possible exposure history of model 1, in which the site has been cyclically covered and uncovered by ice for long enough that nuclide concentrations have reached a dynamic steady state, in which nuclide production during interglacial periods when the site is exposed is equal to radioactive decay of the existing nuclide inventory during each cycle (henceforth 'model 2'). We assume 350 g cm^{-2} cover and zero erosion as discussed above, and we assume 100-kyr durations for glacial–interglacial cycles, so model 2 has one free parameter: the length of time during each cycle that the site is ice-free. Palaeoclimate data, of course, indicate that 100-kyr glacial–interglacial periodicity was not evident before about 0.8 Myr ago, so model 2 is only an approximate representation of palaeoclimate data. Instead, it represents a simple, but infinitely long, exposure history as a contrast to the shortest possible exposure history that can explain the data (model 1). Interestingly, even though model 2 has only one fitting parameter (the fraction of each 100-kyr cycle during which the site is ice-free) we can obtain a good fit to the measurements (Extended Data Fig. 4). The best-fitting interglacial duration is 8,000 years, which fits the data with reduced $\chi^2 = 1.75$ for 6 degrees of freedom, that is, a probability-of-fit of 0.11, which is indistinguishable from the probability of fit for model 1.

We have now discussed two endmember models that fit the data: a single period of middle Pleistocene exposure followed by 1.1 Myr of continuous burial (model 1; this is the shortest possible exposure model that can fit the data), and a dynamic steady state model in which the GIS is absent for 8 kyr of each 100-kyr-long glacial–interglacial cycle (Model 2; as this model is infinitely long, it is the longest possible exposure model that can fit the data, although it is not uniquely so). There are many intermediate models that also fit the data. These are constrained by the general principle that if we start with model 1 and add any periods of exposure during the most recent 1.1 Myr, then we must also add periods of burial before 1.1 Myr ago to maintain the observed $^{26}\text{Al}/^{10}\text{Be}$ ratio. We show several such models, one of which is shown in Fig. 3 as model 3 (see also Extended Data Fig. 4). All these models have the same probability of fit, which is also the same as for model 1 above (and, in fact, they predict exactly the same ^{26}Al and ^{10}Be concentrations).

To summarize, we present endmember exposure models and a wide variety of exposure models between those endmembers that can explain the observed ^{10}Be and ^{26}Al concentrations in the GISP2 bedrock core. The important point of these calculations is not that any one of these scenarios is either uniquely most likely or most favoured by other palaeoclimate data, but that exposure histories that include ice-free conditions during any or all Pleistocene interglaciations (before the present one) can be fitted to the data. However, a scenario in which the present ice sheet has existed continuously for more than the last 1.1 Myr does not fit the observations under any circumstances.

Code availability. MATLAB code used to perform all data analysis in this paper is available at <http://hess.ess.washington.edu/repository/GISP2>.

Erosion of bedrock under the GIS summit. First, the observation by Bierman and others⁴ of high concentrations of meteoric ^{10}Be in soil within GISP2 basal ice is not consistent with any substantial erosion of the underlying bedrock. Second, available data indicate that erosion from normal ice-sheet processes is very slow, and inadequate to remove a substantial amount of bedrock underneath the GIS summit. Summarized by ref. 25, the basal $h_b \approx 13$ m of ice in the GISP2 ice core contained a volume-concentration of silt $c_s \approx 0.001$. Measured borehole deformation indicates a mean horizontal flow velocity for this silty ice of $u \approx 0.07 \text{ m yr}^{-1}$. If the rock material is sourced uniformly from the $L \approx 2.8 \times 10^4 \text{ m}$ flow line extending through GISP2 from the ice divide at GRIP, then the steady-state erosion rate is $E = h_b c_s u / L \approx 3 \times 10^{-8} \text{ m yr}^{-1}$ of rock, or 3 cm per million years²⁵.

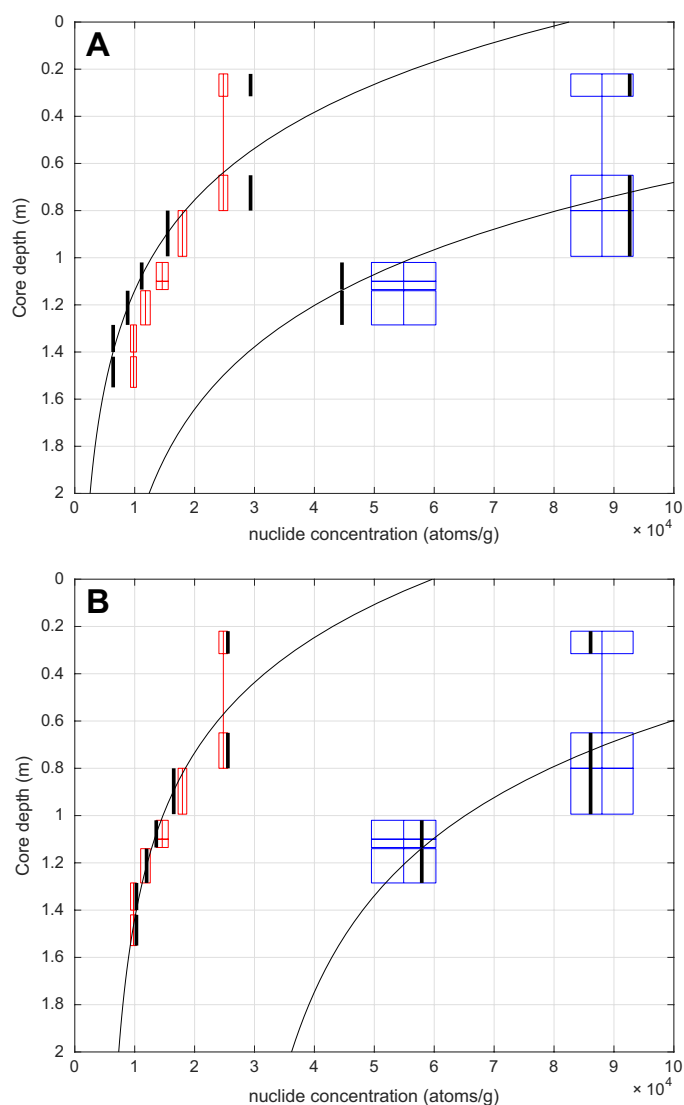
Basal silty ice chronologies and bedrock exposure ages. The new results of Yau *et al.*¹⁰ on the antiquity of basal ice from the Dye-3 (southern Greenland) and GRIP (summit) ice cores might at first seem to raise questions about the results presented here, but may indeed provide a consistent history and point to ways to constrain that history more tightly. Using the measured ratios of argon isotopes in air trapped in the basal ice of the cores, together with the rise in atmospheric ^{40}Ar over time linked to radioactive decay of ^{40}K in rocks and subsequent degassing, the oldest ages of replicates of air that Yau *et al.*¹⁰ found in ice at the base of the GRIP core gave $970 \pm 140 \text{ kyr}$ (and $400 \pm 170 \text{ kyr}$ in the Dye-3 core); much younger air was also present. In the light of the possibility of contamination by local radiogenic production, the ice containing the oldest air could be even older.

These findings are not themselves inconsistent with our data, because our main conclusion is prolonged periods of nearly ice-free Greenland during the Pleistocene, and the simplest GIS scenario consistent with our data that also gives an upper bound for the maximum period the GIS was continuously present (the maximum-stability scenario '1' in Fig. 3) is that of approximately 280 kyr of ice-free Greenland summit before 1.1 Myr ago followed by burial by the GIS since then. Alternatively, it could be that the old basal ice at GRIP and Dye-3 was preserved during periods of nearly vanished GIS in the eastern or northeastern Greenland highlands and subsequently flowed to their current positions. Modelling of the GIS under warming, or under cooling from warmer conditions, often produces a greatly reduced ice volume together with ice-free conditions across most of the island including at or near GISP2, but an intact small ice cap on the highlands to the east (for example, see ref. 27); the simulated position of the western boundary of that ice cap relative to GISP2 and GRIP (which is 28 km east of GISP2) depends on the model, including its resolution, and on the forcing, history, and so on (Fig. 1). The broad picture of retreat to and advance from a greatly reduced eastern

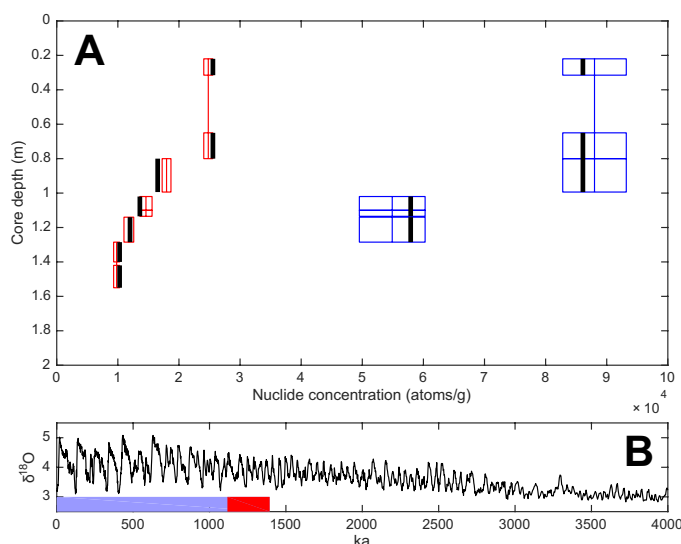
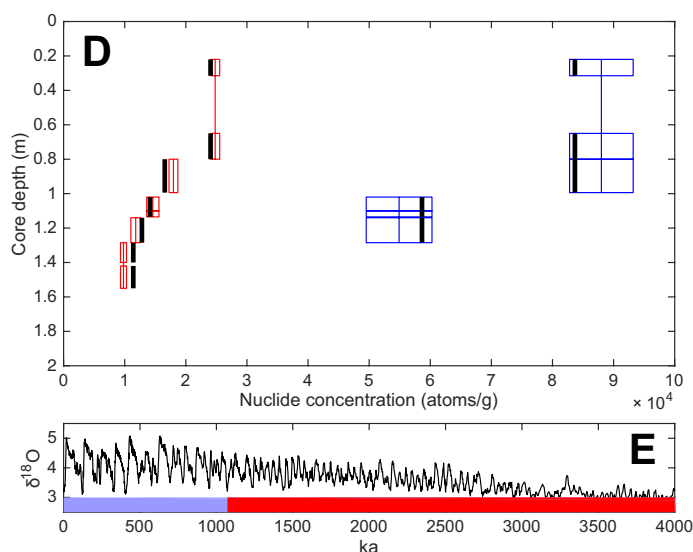
ice cap is fairly widespread across models, and histories can be simulated in which substantial ice-sheet shrinkage and regrowth occur with deglaciation of GISP2 and GRIP locations, or with deglaciation of GISP2 but not of GRIP. Notably, if advance occurs from an extant ice cap over adjacent surfaces, old ice will move over younger surfaces. Thus, our results plus those of Yau *et al.*¹⁰ are consistent with the idea of preservation of old ice in the eastern highland areas during prominent interglacials, as discussed in the main text. The bedrock cosmogenic nuclides and the basal gas isotopes provide independent, potentially complementary constraints, and together might deliver a rather tight framework for ice sheet models.

Data availability. The data used in this study are available in Extended Data Tables 1–3.

31. Schaefer, J. M. *et al.* High frequency Holocene glacier fluctuations in New Zealand differ from the northern signature. *Science* **324**, 622–625 (2009).
32. Davis, J. C. *et al.* LLNL/UC AMS Facility and Research-Program. *Nucl. Instrum. Methods Phys. Res. B* **52**, 269–272 (1990).
33. Granger, D. E. *et al.* New cosmogenic burial ages for Sterkfontein Member 2 Australopithecus and Member 5 Oldowan. *Nature* **522**, 85–88 (2015).
34. Paul, M. Separation of isobars with a gas-filled magnet. *Nucl. Instrum. Methods Phys. Res. B* **52**, 315–321 (1990).
35. Timmers, H., Weijers, T. D. M. & Elliman, R. G. Unique capabilities of heavy ion elastic recoil detection with gas ionization detectors. *Nucl. Instrum. Methods Phys. Res. B* **190**, 393–396 (2002).
36. Fifield, L. K., Tims, S. G., Gladkiss, L. G. & Morton, C. R. ^{26}Al measurements with ^{10}Be counting statistics. *Nucl. Instrum. Methods Phys. Res. B* **259**, 178–183 (2007).
37. Balco, G., Stone, J. O., Lifton, N. A. & Dunai, T. J. A complete and easily accessible means of calculating surface exposure ages or erosion rates from Be -10 and Al -26 measurements. *Quat. Geochronol.* **3**, 174–195 (2008).
38. Stone, J. Air pressure and cosmogenic isotope production. *J. Geophys. Res.* **105**, 23753–23759 (2000).
39. Nishiizumi, K. Preparation of Al -26 AMS standards. *Nucl. Instrum. Methods Phys. Res. B* **223/224**, 388–392 (2004).
40. Borchers, B. *et al.* Geological calibration of spallation production rates in the CRONUS-Earth project. *Quat. Geochronol.* **31**, 188–198 (2016).
41. Phillips, F. M. *et al.* The CRONUS-Earth project: a synthesis. *Quat. Geochronol.* **31**, 119–154 (2016).
42. Rovey, C. W. & Balco, G. Paleoclimatic interpretations of buried paleosols within the pre-Illinoian till sequence in northern Missouri, USA. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **417**, 44–56 (2015).
43. Bierman, P. R., Marsella, K. A., Patterson, C., Davis, P. T. & Caffee, M. Mid-Pleistocene cosmogenic minimum-age limits for pre-Wisconsinan glacial surfaces in southwestern Minnesota and southern Baffin island: a multiple nuclide approach. *Geomorphology* **27**, 25–39 (1999).

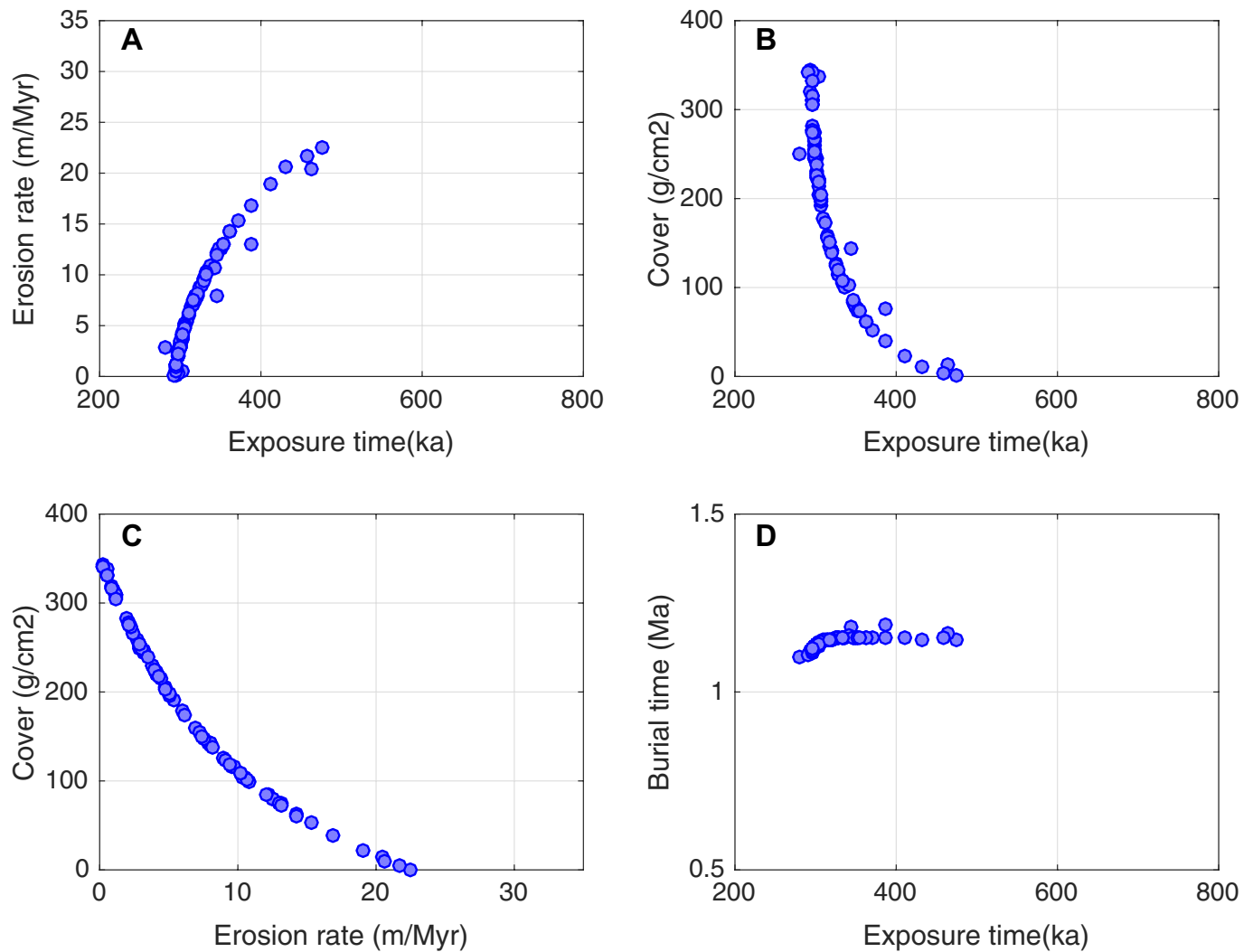


Extended Data Figure 1 | ^{10}Be and ^{26}Al concentrations compared to production rates. ^{10}Be ($n = 5$) and ^{26}Al ($n = 2$) concentration–depth profiles compared to the depth dependence of nuclide production rates (including both spallogenic and muon production). In both **a** and **b**, the red and blue boxes represent ^{10}Be (red) and ^{26}Al (blue) measurements. The vertical dimension of each box represents a distinct segment of core, and vertical lines connect multiple core segments that were amalgamated for each ^{10}Be or ^{26}Al analysis. In most cases amalgamated segments were adjacent to each other, but in some cases (for example, the uppermost two core segments), they were separated by gaps. The width of the boxes shows measurement uncertainty (1σ ; see Extended Data Table 1) on nuclide concentrations. In **a**, we attempt to fit the observations by assuming that the bedrock surface is the land surface, the erosion rate is zero, and by allowing the duration of a single period of exposure to vary. For this exercise we treat ^{10}Be and ^{26}Al separately, that is, the predictions are not forced to obey the production ratio. The continuous thin black lines show predicted nuclide concentrations for this model, and the discontinuous, darker, black bars show predicted nuclide concentrations averaged over depth ranges corresponding to each analysis. The black bars, therefore, are the model predictions that we compare to the measurements. This model cannot be fitted to the data, because observed nuclide concentrations do not decrease as rapidly as they would if the bedrock surface were the land surface during exposure. In **b**, we perform the same fitting exercise, but also include a thickness of shielding mass above the bedrock surface as an additional fitting parameter. This removes the systematic misfit shown in **a** and makes it possible to fit the observations. A cover thickness of 350 g cm^{-2} best fits the observations.

A-C: Model 1**D-F: Model 1B**

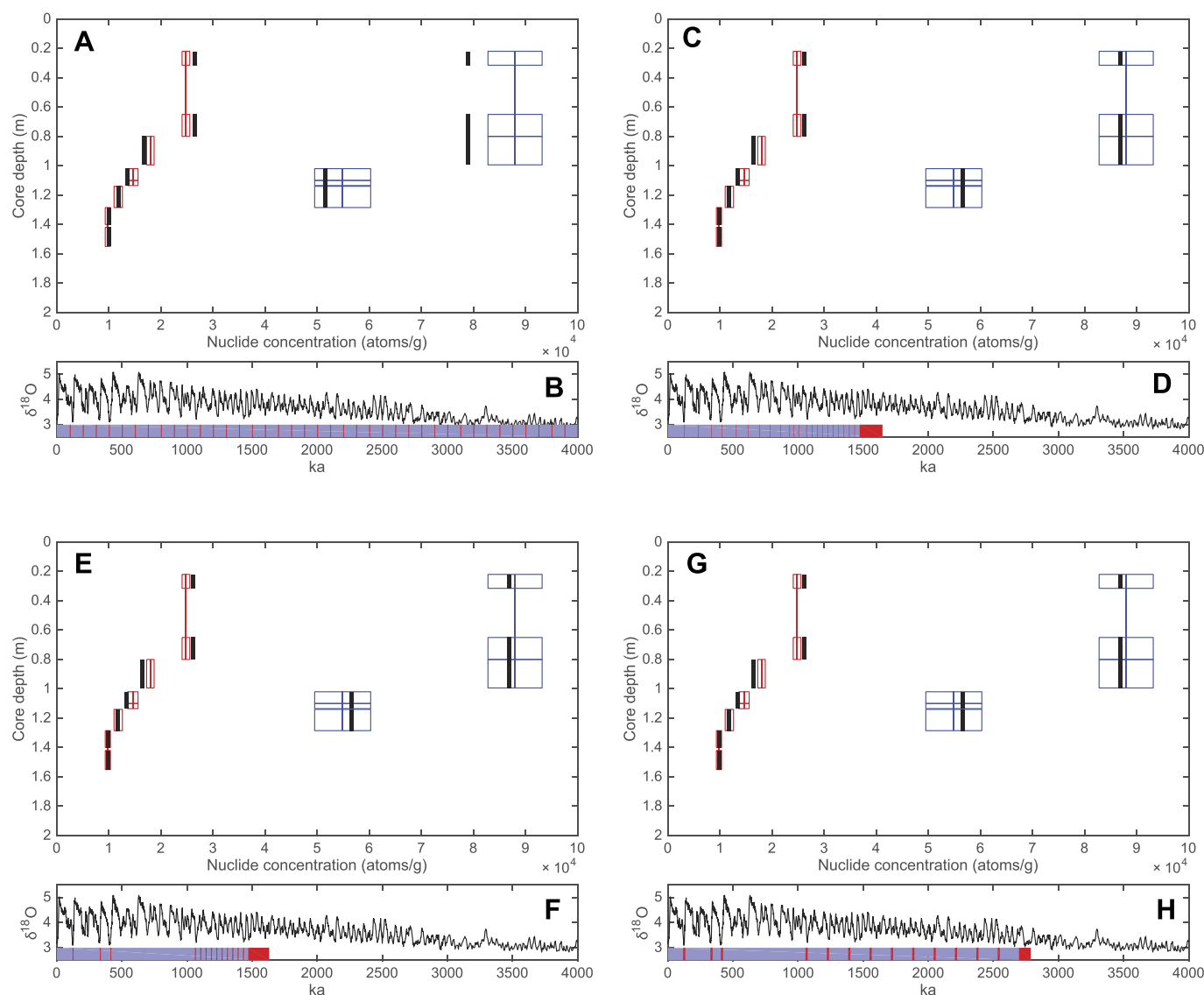
Extended Data Figure 2 | Two-stage exposure histories fit to ^{10}Be and ^{26}Al measurements. In **a** and **d**, the red and blue boxes represent ^{10}Be ($n=5$; red) and ^{26}Al ($n=2$; blue) measurements. The vertical dimension of each box represents a distinct segment of core, and vertical lines connect multiple core segments that were amalgamated for each ^{10}Be or ^{26}Al analysis. In most cases amalgamated segments were adjacent to each other, but in some cases (for example, the uppermost two core segments), they were separated by gaps. The width of the boxes shows measurement uncertainty (1σ ; see Extended Data Tables 2 and 3) on nuclide concentrations. The thick black lines are nuclide concentrations predicted for each core segment by the best-fitting parameters of each model. **b** and **e** show the exposure history implied by the best-fitting parameters of each model compared to the LR04 oxygen isotope stack from ref. 30. Red bars represent periods of surface exposure and blue bars periods of cover by the ice sheet. **c** and **f** show observed nuclide concentrations compared to model predictions for samples in which both ^{10}Be and ^{26}Al were measured, normalized to production rates at their respective depths implied by each model, on a two-nuclide diagram²⁹. Red ellipses are 68% confidence regions for the nuclide concentrations

including measurement uncertainties only, and black dots are nuclide concentrations predicted by best-fitting model parameters. The solid black lines show the simple exposure region; darker dashed lines are isolines of burial in increments of 1 Myr, and lighter dotted lines are isolines of exposure time in increments of 0.1 Myr. **a**, **b** and **c** show the fit of model 1, the simplest possible model that fits the data, which includes a single period of surface exposure, a single period of burial, 350 g cm^{-2} of cover thickness above the bedrock surface during exposure, and zero surface erosion. In **c**, nuclide concentrations are normalized to production rates at sample depths below this additional cover thickness. This model provides a good fit to the measurements. The exposure history implied by the best-fitting parameters for model 1 (280-kyr exposure, 1.1 Myr burial) provides a maximum limiting constraint on the length of time the present ice sheet has been continuously present at the core site. **d**, **e** and **f** show the fit of model 1B, which includes a long period of continuous exposure with steady surface erosion, a single period of burial, and zero additional cover thickness above the bedrock surface. In **f**, nuclide concentrations are normalized to production rates at sample depths below the bedrock surface. This model cannot be adequately fitted to the observations.



Extended Data Figure 3 | Two-stage exposure-burial models with four free parameters. a–d show parameter values for two-stage exposure-burial models with four free parameters (model 1C: free parameters are exposure time t_{exp} , burial time, surface erosion rate during exposure, and additional cover thickness above the bedrock surface during exposure

periods) that yield acceptable fits to the observations. All panels show the same set of results; only the axes differ. Note that model 1 with best-fitting parameters (280-kyr exposure, 1.1 Myr burial, zero surface erosion, 350 g cm⁻² cover thickness) is an endmember of this distribution.



Extended Data Figure 4 | Many-stage exposure histories fitted to ^{10}Be and ^{26}Al measurements. In **a**, **c**, **e** and **f**, the red and blue boxes represent ^{10}Be (red) and ^{26}Al (blue) measurements. The vertical dimension of each box represents a distinct segment of core, and vertical lines connect multiple core segments that were amalgamated for each ^{10}Be or ^{26}Al analysis. In most cases amalgamated segments were adjacent to each other, but in some cases (for example, the uppermost two core segments), they were separated by gaps. The width of the boxes shows measurement uncertainty (1σ ; see Extended Data Tables 2 and 3) on nuclide concentrations. The thick black lines are nuclide concentrations predicted for each core segment by the best-fitting parameters of each model. **b**, **d**, **f** and **h** show the exposure history implied by the best-fitting parameters of each model compared to the LR04 oxygen isotope stack from ref. 30. Red bars represent periods of surface exposure and blue bars represent periods of cover by the ice sheet. **a** and **b** show model 2, which is a dynamic steady state model with 100-kyr cycles. This model has one free parameter, which is the fraction of each cycle during which the site is ice-free, and assumes zero surface erosion and 350 g cm^{-2} cover thickness above the bedrock surface. The best-fitting length of ice-free periods

for this model is 8,000 years. All other panels show models with ice-free conditions during some or all middle and late Pleistocene interglaciations, and have two free parameters: the duration of ice-free conditions during interglaciations, and the length of an initial ice-free period in the middle or early Pleistocene. **c** and **d** show a model in which the core site is ice-free during all interglacials within the period of 100-kyr-long glacial–interglacial cycles after 1.1 Myr ago, during an (arbitrarily long) series of short interglacials during 41-kyr-long glacial–interglacial cycles before 1.1 Myr ago, and during an initial longer period of exposure in the middle Pleistocene. The best-fitting duration of ice-free conditions during interglacials for this model is 4,200 years. **e** and **f** show ice-free conditions during MIS 9, MIS 11 and MIS 13, an (arbitrarily long) period of short ice-free interglaciations during 41-kyr-long cycles before 1.1 Myr ago, and an initial period of continuous exposure in the middle Pleistocene. The best-fitting duration of ice-free conditions during interglacials for this model is 7,400 years. **g** and **h** show ice-free conditions during MIS 9, MIS 11 and MIS 13, with (arbitrarily located) occasional ice-free periods in the early and middle Pleistocene. The best-fitting duration of ice-free conditions during interglacials for this model is 18,200 years.

Extended Data Table 1 | ^{10}Be and ^{26}Al concentrations of the GISP2 bedrock core

	Depth Top (m)	Depth Bottom (m)	mid-inventory depth (m)	Quartz (g)	^{10}Be (at/g)	1 sigma	^{26}Al (at/g)	1sigma	$^{26}\text{Al}/^{10}\text{Be}$	1 sigma	% error
GISP 2/3	0.22	0.65	0.44	12.66	24,800	740					
GAP	0.32	0.65									
GISP 4	0.80	1.00	0.89	17.88	18,000	720					
GISP 2/3/4	0.22	1.00	0.49	30.54	20,800	520	88,000	5,200	4.23	0.27	6.4%
GAP	1.00	1.02									
GISP 5/6	1.02	1.14	1.07	9.34	14,600	1,000					
GISP 7	1.14	1.29	1.21	8.81	11,800	800					
GISP 5/6/7	1.02	1.29	1.14	18.15	13,300	1,300	55,000	5,400	4.14	0.57	13.9%
GISP 8/9	1.29	1.55	1.40	8.87	9,800	490					

^{10}Be ($n = 5$) and ^{26}Al ($n = 2$) depth profile data from the GISP-2 core. The mid-inventory depth is the depth dividing each core piece into two halves of equal ^{10}Be and ^{26}Al inventory. See also core stratigraphy in Fig. 1. We measured 5 core samples for ^{10}Be . We combined the Al fractions (and splits for ^{27}Al analysis) of GISP2/3 and GISP4 as well as of GISP5/6 and GISP7 to get two precise $^{26}\text{Al}/^{27}\text{Al}$ measurements. The given 1σ errors reflect analytical uncertainties. To compare ^{10}Be and ^{26}Al measurements directly in the same sample and to calculate the $^{26}\text{Al}/^{10}\text{Be}$ ratio, we also show the ^{10}Be for GISP2/3/4 and GISP5/6/7, which are simply the sums of the measured ^{10}Be atoms in the respective ^{10}Be samples (GISP2/3 + GISP4 and GISP5/6 + GISP7). The 'Quartz' column gives the sample size.

Extended Data Table 2 | GISP2 bedrock ^{10}Be data ^{10}Be

Sample	Quartz (g)	^9Be carrier (mg)	$^{10}\text{Be}/^9\text{Be}$ (10^{-15})	1σ (10^{-15})	Uncertainty %	^{10}Be atoms	1σ error	^{10}Be (atoms/g)	1σ error	Max ^{10}Be Current (microAmps)
BLK1-2015Jul22			0.32	0.12	38%	3900	150			23
BLK2-2015Jul22			0.45	0.12	27%	5500	150			19
BLK3-2015Jul22			0.43	0.11	26%	5300	140			18
BLK4-2015Jul22			0.13	0.17	131%	1600	2100			20
BLK5-2015Jul22			0.39	0.12	31%	6000	190			17
GISP 2/3	12.66	0.1889	25.20	0.82	3%	318,200	9,600			21
<u>GISP 2/3 corrected</u>	<u>12.66</u>		<u>24.86</u>	<u>0.83</u>	<u>3%</u>	<u>313,500</u>	<u>9,405</u>	<u>24,760</u>	<u>740</u>	
GISP 4	17.88	0.1894	25.72	0.96	4%	325,700	13,000			18
<u>GISP 4 corrected</u>	<u>17.88</u>		<u>25.38</u>	<u>0.97</u>	<u>4%</u>	<u>321,300</u>	<u>12,852</u>	<u>17,970</u>	<u>720</u>	
<u>GISP 2/3/4</u>	<u>30.54</u>					<u>634,800</u>	<u>15,926</u>	<u>20,790</u>	<u>520</u>	
CB_2015May			0.05	0.10	200%	600	1200			26
BLK1-2015Apr17			0.38	0.09	24%	4700	1100			23
BLK2-2015Apr17			0.58	0.24	41%	7100	2900			20
BLK3-2015Apr17			0.01	0.22	2200%	100	2200			21
BLK4-2015Apr17			0.33	0.09	27%	4100	1100			21
BLK1-2015Apr21			0.3	0.09	30%	4000	1200			30
GISP 5/6	9.34	0.1888	11.10	0.65	6%	140,000	8,400			25
<u>GISP 5/6 corrected</u>	<u>9.34</u>		<u>10.80</u>	<u>0.71</u>	<u>7%</u>	<u>136,300</u>	<u>9,500</u>	<u>14,600</u>	<u>1,020</u>	
GISP 7	8.81	0.1886	8.57	0.46	5%	108,000	5,400			25
<u>GISP 7 corrected</u>	<u>8.81</u>		<u>8.26</u>	<u>0.54</u>	<u>7%</u>	<u>104,100</u>	<u>7,300</u>	<u>11,800</u>	<u>830</u>	
<u>GISP 5/6/7</u>	<u>18.15</u>					<u>240,400</u>	<u>11,981</u>	<u>13,245</u>	<u>660</u>	
CB1_2014Apr23			0.05	0.25	500%	600	3000			25
CB2_2014Apr23			0.05	0.12	240%	600	1400			23
Blk-4_2014Feb28			0.30	0.11	37%	4000	1500			16
Blk-5_2014Feb28			0.23	0.39	170%	2800	4800			14
Blk-6_2014Feb28			0.19	0.49	260%	2300	6000			15
GISP 8/9	8.87	0.1449	9.21	0.49	5%	89,200	4,500			17
<u>GISP 8/9 corrected</u>	<u>8.87</u>		<u>8.97</u>	<u>0.49</u>	<u>5%</u>	<u>86,900</u>	<u>4,400</u>	<u>9,800</u>	<u>490</u>	

Sample and measurement details for the ^{10}Be samples, given with 1σ analytical uncertainty. Underlined values are the measurements used for further discussion. 'BLK' denotes process blanks, that is, ^9Be carrier that underwent a chemical procedure identical to that of the respective samples. We processed 3–5 such process blanks for each sample to ensure robust monitoring of any contamination; the background correction was done by subtracting the mean of the ^{10}Be atoms for all BLKs from the total ^{10}Be atoms in the corresponding sample. 'CB' denotes carrier blanks, that is, ^9Be carrier that did not undergo the sample and process blank chemistry, but was precipitated, dried, loaded in cathodes and sent directly to the Center for Accelerator Mass Spectrometry at the Lawrence Livermore National Laboratory for $^9\text{Be}/^{10}\text{Be}$ analysis. The difference between 'CB' and 'BLK' ^{10}Be atoms allows us to estimate the ^{10}Be contamination during the sample processing in the laboratory, which is small for the samples discussed here. During the 23 years of storage at the National Ice Core Laboratory in Boulder (1,600 m), the potential production of ^{10}Be and ^{26}Al in the core is of the order of several hundred atoms and is thus negligible compared to the measured values.

Extended Data Table 3 | GISP2 bedrock ^{26}Al data

²⁶ Al														
Sample	Quartz (g)	total ²⁷ Al (mg)	natural ²⁷ Al (mg)	²⁶ Al/ ²⁷ Al (%) (10 ⁻¹⁵)	1 σ (10 ⁻¹⁵)	Uncertainty %	²⁶ Al (atoms)	1 σ error	²⁶ Al (atoms/g)	1 σ error	Max ²⁶ Al Current (microAmps)	Min ²⁶ Al Current (microAmps)	Total ²⁶ Al counts	Measurement time (min)
BLK5-2015Jul22				0.48	0.48	99.0%					11	8	2	
BLK3/4-2015Jul22				0.55	0.52	95.8%					10	9	2	
BLK2-2015Jul22				0.27	0.39	144.1%					10	8	1	
GISP-2/3/4	30.54	1.9117	0.6872	64.12	3.70	5.8%					11	7	351	35
GISP-2/3/4 corrected ¹				63.69	3.70	5.8%								
GISP-2/3/4 corrected ²				62.99	3.72	5.9%	2,687,976	158,744	88,015	5,198				
BLK 5/6/7-2015Apr17				2.55	0.71	27.9%					8	4	13	
BLK2-2015Apr17				5.50	1.95	35.5%					9	4		
GISP-5/6/7	18.15	1.8232	0.3648	28.13	2.48	8.8%					8	4	143	55
GISP-5/6/7 corrected ¹				24.10	3.24	13.4%								
GISP-5/6/7 corrected ²				24.49	2.40	9.8%	996,688	97,675	54,914	5,382				
BLK1-2015Mar14				2.48	0.82	33.3%					12	2	12	
BLK2-2015Mar14				1.69	0.64	37.7%					11	1	11	
GISP 8/9	7.56	1.7503		8.39	0.91	10.9%					12	1	92	100

Sample and measurement details for the ^{26}Al samples. 'BLK' denotes process blanks, that is, ^{27}Al carrier that underwent a chemical procedure identical to that of the respective samples. Natural ^{27}Al was determined from two independent splits taken from the fully digested quartz sample by ICP-OES analysis, calibrated with in-house, gravimetrically prepared ^{27}Al standards cross-checked with commercial standards.

¹The mean of the corresponding BLK values was subtracted from the sample to give the corrected value.

²To better account for memory effects during the AMS measurements, we performed a time-progressive background correction: $^{26}\text{Al}/^{27}\text{Al}$ measurements at PRIME laboratory are done in one-minute cycles and we directly subtracted BLK counts (counts per minute) closest to the measurement cycle of the corresponding sample.

Both background corrections yield very similar results, and we use the data based on the time-progressive background correction. See Methods for additional details. GISP2/3/4 and corresponding BLKs were analysed with AMS standards at PRIME mixed with niobium, while the other samples were measured versus AMS standards mixed with silver. The ^{26}Al measurement of the lowest sample, GISP8/9, given in small italics, was an initial sensitivity test: we added ^{27}Al carrier to the Al fraction after the cation column separation, so that any aluminium loss during the geochemical processing would go undetected; also, the 'count-rate-based blank correction' for this measurement was substantial and the measured $^{26}\text{Al}/^{27}\text{Al}$ ratio fell into the 10^{-15} range, which is near the detection limit of the AMS analysis and thus challenging; taken together, the ^{26}Al measurement of GISP8/9 is less reliable and not directly comparable to those of GISP5/6/7 and GISP2/3/4, and we base our discussion on these two upper samples.

A persistent and dynamic East Greenland Ice Sheet over the past 7.5 million years

Paul R. Bierman¹, Jeremy D. Shakun², Lee B. Corbett¹, Susan R. Zimmerman³ & Dylan H. Rood^{4,5}

Climate models show that ice-sheet melt will dominate sea-level rise over the coming centuries, but our understanding of ice-sheet variations before the last interglacial 125,000 years ago remains fragmentary. This is because terrestrial deposits of ancient glacial and interglacial periods^{1–3} are overrun and eroded by more recent glacial advances, and are therefore usually rare, isolated and poorly dated⁴. In contrast, material shed almost continuously from continents is preserved as marine sediment that can be analysed to infer the time-varying state of major ice sheets. Here we show that the East Greenland Ice Sheet existed over the past 7.5 million years, as indicated by beryllium and aluminium isotopes (¹⁰Be and ²⁶Al) in quartz sand removed by deep, ongoing glacial erosion on land and deposited offshore in the marine sedimentary record^{5,6}. During the early Pleistocene epoch, ice cover in East Greenland was dynamic; in contrast, East Greenland was mostly ice-covered during the mid-to-late Pleistocene. The isotope record we present is consistent with distinct signatures of changes in ice sheet behaviour coincident with major climate transitions. Although our data are continuous, they are from low-deposition-rate sites and sourced only from East Greenland. Consequently, the signal of extensive deglaciation during short, intense interglacials could be missed or blurred, and we cannot distinguish between a remnant ice sheet in the East Greenland highlands and a diminished continent-wide ice sheet. A clearer constraint on the behaviour of the ice sheet during past and, ultimately, future interglacial warmth could be produced by ¹⁰Be and ²⁶Al records from a coring site with a higher deposition rate. Nonetheless, our analysis challenges the possibility of complete and extended deglaciation over the past several million years.

Understanding of early Greenlandic glaciation remains fragmentary, uncertain and for some periods contradictory (Fig. 1 and references therein); much of what is known comes from marine sediment. The first presence of ice-rafted debris (IRD) at Ocean Drilling Program (ODP) site 918 suggests that East Greenland glaciers initially reached the coast about 7.5 million years (Myr) ago⁶, whereas the surface texture of quartz at Site 918 suggests that glaciation there began 11 Myr ago⁷. IRD data suggest that the first large-scale glaciation of Greenland occurred in the latest Pliocene (3.3 Myr ago), and multiple IRD records indicate expansive Greenland glaciation by 2.7 Myr ago⁸. Poorly dated shallow marine deposits indicate periods of warmth in Greenland during the later Pliocene or early Pleistocene, some after the initial onset of glaciation⁴. Geochemical and pollen data suggest that southern Greenland was at least partially deglaciated and forested during Marine Isotope Stage (MIS) 11 (about 400,000 years (400 kyr) ago) and perhaps MIS 5e (about 130 kyr ago). In contrast, meteoric ¹⁰Be data from silt at the base of the Greenland Ice Sheet Project 2 ice core (GISP2 in east-central Greenland) are consistent with continuous cover there by cold-based, non-erosive ice for millions of years⁹, an assertion supported by noble gas measurements¹⁰.

The concentration of cosmogenic nuclides in Earth materials reveals near-surface history^{11,12}. In non-glaciated terrain, cosmic rays bombard Earth and produce ¹⁰Be and ²⁶Al in mineral lattices. Production rates and nuclide concentrations decrease exponentially within a few metres of the surface; weak muon interactions continue at lower rates for tens of metres¹³ (Fig. 2a). Covering a landscape with ice stops cosmogenic-nuclide production in the underlying rock. Subsequent glacial erosion first removes the most highly irradiated, near-surface material before excavating rock from depths containing progressively lower isotope concentrations (Fig. 2b).

Thermal conditions at the ice-sheet bed control its ability to erode, incorporate and transport rock and sediment. Warm-based ice (at the pressure melting point) can effectively erode rock and transport sediment to and off the coast¹⁴; thus, the isotopic record we present here is strongly biased towards areas of the ice sheet that were warm-based¹⁵. Cold-based ice, below the pressure melting point, is frozen to the bed and generally non-erosive¹⁶; it buries and preserves ancient landscapes rather than eroding them.

The ratio of the cosmogenic nuclides ¹⁰Be and ²⁶Al provides additional information about burial after initial exposure¹². Because ²⁶Al (half-life $t_{1/2} = 0.71$ Myr) radiodecays more rapidly than ¹⁰Be ($t_{1/2} = 1.39$ Myr), burial of previously exposed material will, over time, lower both the ²⁶Al/¹⁰Be ratio and the concentration of both isotopes (Fig. 2b). ²⁶Al and ¹⁰Be are produced at a ratio of 7.3 ± 0.3 (1 σ) near sea level and at high latitude in Greenland (Methods); thus, measured ²⁶Al/¹⁰Be ratios falling below 7.3 are diagnostic of burial for long periods of time (more than about 200 kyr).

We measured cosmogenic-nuclide records spanning the last 7.5 Myr and 2.2 Myr at ODP sites 918 and 987, respectively (hereafter Site 918 and Site 987; Fig. 3; Extended Data Figs 1 and 2). Starting in the Miocene at Site 918, decay-corrected ¹⁰Be concentrations generally decrease as sediment gets younger, reflecting the progressive glacial erosion of once-stable regolith and bedrock in East Greenland (Fig. 4c). Concentrations of ¹⁰Be, corrected for radiodecay on the seafloor, are high (135,000 atoms per gram of quartz sediment) in the oldest glacial sediment⁶ (about 7.5 Myr old; sample 918-30) and indicate that the pre-glacial East Greenland landscape was eroding at approximately 22 m per million years (Methods). As sediment and rock were removed from the landscape by glacial erosion, material that was deeply shielded in pre-glacial times, and thus less affected by cosmic radiation, was incorporated into basal ice and carried offshore. We infer that the decrease in ¹⁰Be concentration was driven by glacial erosion because IRD at Site 918 and at other sites in East Greenland^{2,5} indicates the presence of glaciers eroding rock, extending to the sea, and supplying the sand-sized sediment that we analysed⁶. A general increase in the intensity and spatial extent of glaciation after 7.5 Myr ago is supported by rising accumulation rates of coarse sediment over time at Site 918 (Fig. 4b)⁶.

By the late Pliocene (around 3 Myr ago), when other records suggest that the first major ice expansion occurred over most of Greenland

¹Department of Geology and Rubenstein School of the Environment and Natural Resources, University of Vermont, Burlington, Vermont 05405, USA. ²Department of Earth and Environmental Sciences, Boston College, Chestnut Hill, Massachusetts 02467, USA. ³Center for Accelerator Mass Spectrometry, Lawrence Livermore National Laboratory, Livermore, California 94550, USA.

⁴Department of Earth Science and Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. ⁵Scottish Universities Environmental Research Centre, East Kilbride G75 0QF, UK.

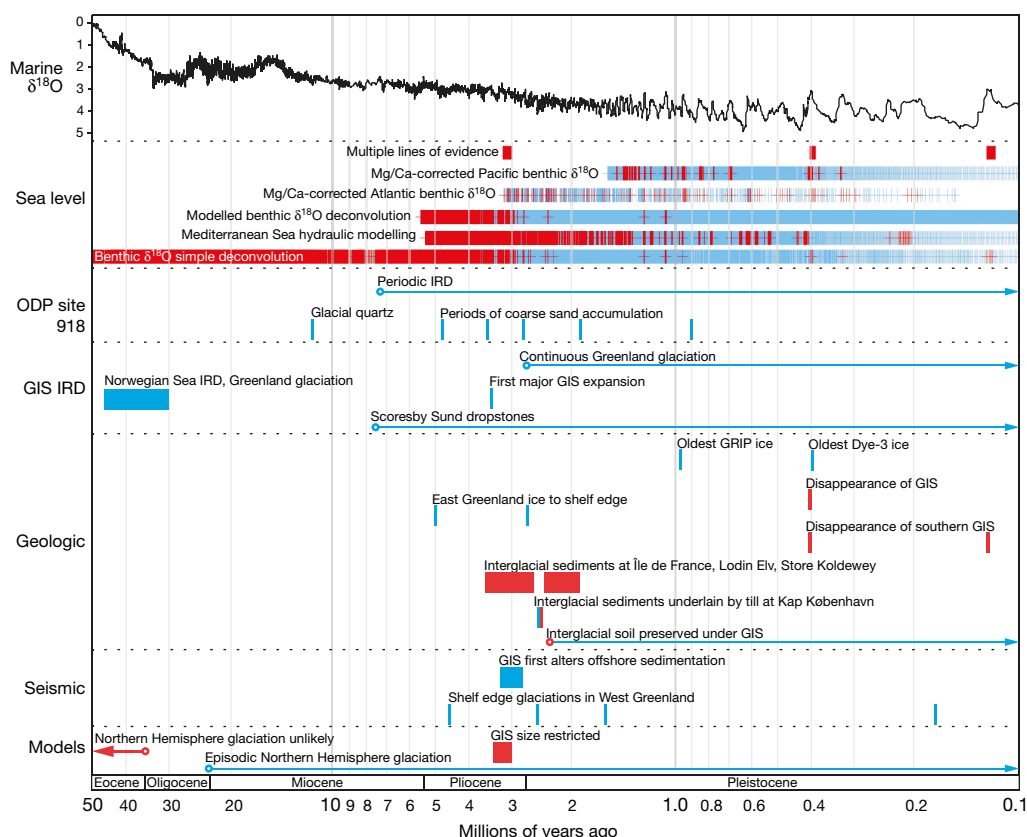


Figure 1 | Compilation of findings that constrain the long-term history of the GIS. Red indicates smaller ice sheet; blue indicates larger ice sheet. References for this figure are included in Extended Data Fig. 5. GRIP, Greenland Ice Core Project; Dye-3, an ice core named after the radar station where the core was collected.

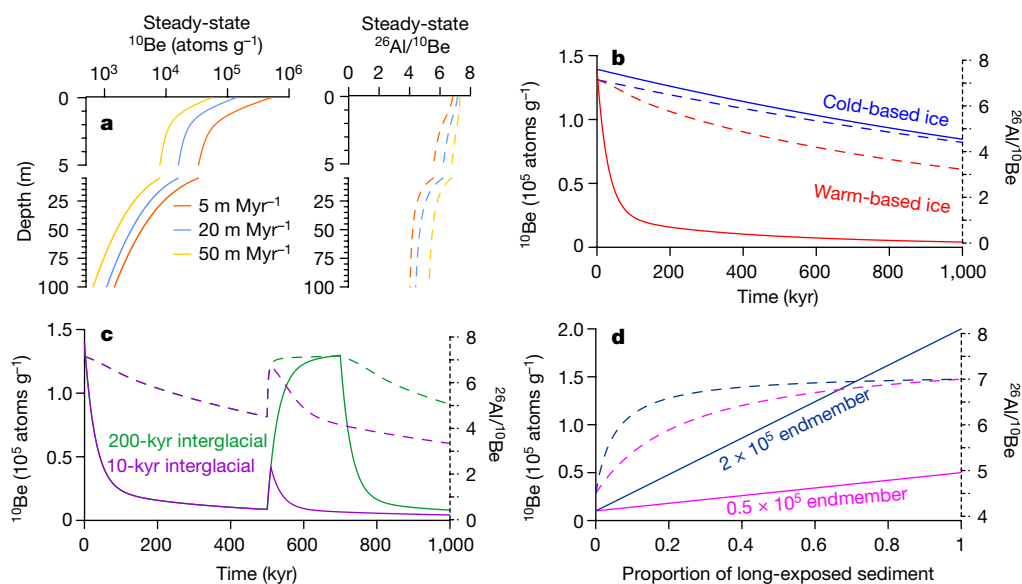


Figure 2 | Cosmogenic-nuclide systematics and sensitivity to erosion, burial, exposure and mixing. **a**, Estimated steady-state concentrations of ^{10}Be and the $^{26}\text{Al}/^{10}\text{Be}$ ratio as a function of depth below the preglacial surface of Greenland (sea level) for various erosion rates. **b**, ^{10}Be concentration (solid lines) and $^{26}\text{Al}/^{10}\text{Be}$ (dashed lines) starting from the 20 m Myr^{-1} steady-state profile following 1 Myr of cold-based ice cover and warm-based ice eroding at 20 m Myr^{-1} . **c**, ^{10}Be concentration (solid lines) and $^{26}\text{Al}/^{10}\text{Be}$ (dashed lines) (starting from the 20 m Myr^{-1} steady-state profile) following 1 Myr of warm-based ice eroding at

20 m Myr^{-1} , but interrupted by either a 10-kyr or 200-kyr interglacial exposure with 20 m Myr^{-1} erosion. **d**, ^{10}Be concentration (solid lines) and $^{26}\text{Al}/^{10}\text{Be}$ (dashed lines) of sediments mixed from varying proportions of an eroded and long-buried endmember ($^{10}\text{Be} = 10,000\text{ atoms g}^{-1}$; $^{26}\text{Al}/^{10}\text{Be} = 4.5$) and two different long-exposed endmembers ($^{10}\text{Be} = 50,000\text{ atoms g}^{-1}$, $200,000\text{ atoms g}^{-1}$; $^{26}\text{Al}/^{10}\text{Be} = 7.3$ in both cases). Muon production rates in all panels were calculated using the MATLAB implementation in refs 13 and 28; see Methods for details.

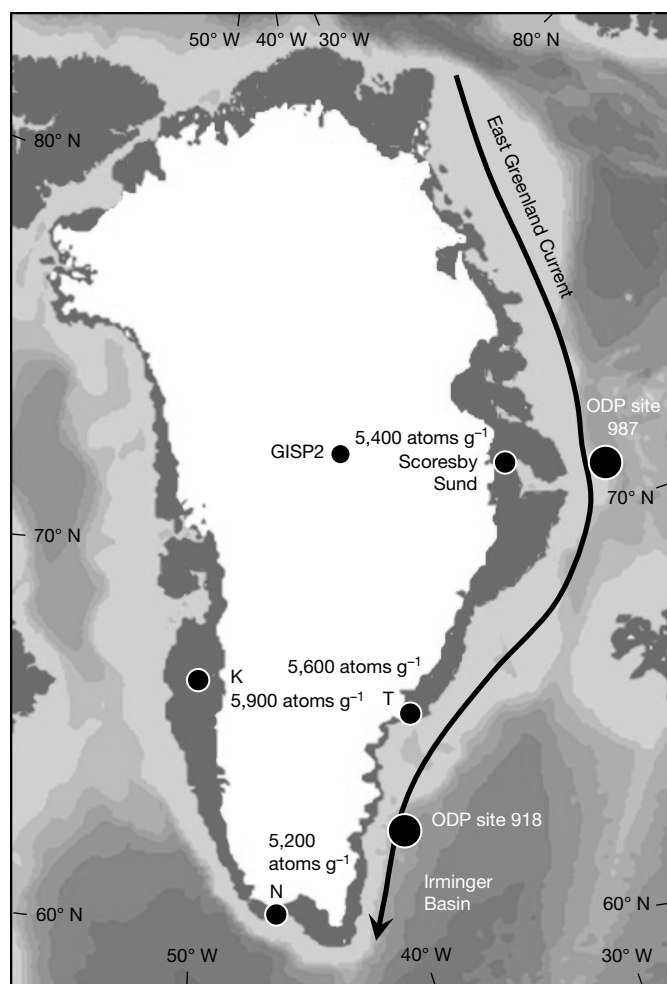


Figure 3 | Map of Greenland. ODP sites used in this study, the East Greenland Current, and locations where contemporary sediment samples were collected are shown. (K, Kangerlussuaq ($n = 26$); N, Narsarsuaq ($n = 16$); T, Tasilaq ($n = 12$); numbers give median ^{10}Be concentrations in these sediments¹⁸ as well as a single value at Scoresby Sund²¹.) Ocean bathymetry contoured in 500-m intervals; derived from MMap (<https://www.eoas.ubc.ca/~rich/map.html>). The locations of the GISP2 ice core site and Irminger Basin are shown.

(Fig. 1), decay-corrected ^{10}Be concentrations are more than an order of magnitude lower than at the beginning of the record, reaching a minimum of 12,000 atoms g^{-1} at 2.8 Myr. These data are consistent with warm-based glaciers having eroded parts of East Greenland for much of the Pliocene, and provide direct evidence for the model- and core-based supposition^{8,17} that the Greenland Ice Sheet (GIS) has been present and eroding East Greenland since at least the Pliocene.

At the dawn of the Pleistocene, decay-corrected ^{10}Be concentration abruptly increases (Fig. 4c). Sediment deposited at around 2.5 Myr ago had about 140,000 atoms g^{-1} of ^{10}Be , more similar to Miocene-age (7.5-Myr-old) sediment than to any of Quaternary age (Supplementary Table 1). This ^{10}Be -rich quartz suggests early Pleistocene expansion of the ice sheet into previously unglaciated areas of East Greenland where stable Miocene regolith remained, an interpretation consistent with abundant IRD found at about 2.5 Myr both at Site 918⁶ and elsewhere in the Arctic⁸. We suspect that this spike in ^{10}Be concentration does not represent an interglacial period, such as that indicated by the warm fauna and flora found in sediment of the Kap København Formation⁴, because the interglacial represented by those sediments is thought to be short-lived (<20 kyr)⁴. Such a short duration is insufficient to raise ^{10}Be concentrations to the levels attained by continuous pre-glacial exposure¹⁸ (Fig. 2c).

From 2.5 Myr to 0.8 Myr ago, the decay-corrected concentration of ^{10}Be generally declines (Fig. 4c), reflecting continued erosion of rock and regolith by warm-based areas of the eastern GIS. The decay-corrected ^{10}Be record at Site 918 shows parallels with core sedimentology (Fig. 4a, b); an overall inverse correlation exists between ^{10}Be concentration and sand content (Extended Data Fig. 3), and several previously noted IRD pulses line up with prominent drops in decay-corrected ^{10}Be concentration, such as pulses near 7 Myr, 2.8 Myr, 1.9 Myr and 0.8 Myr ago (Fig. 4b, c)¹⁹. These patterns are consistent with periods of intensified glacial erosion that excavated deeper-sourced material containing less ^{10}Be and delivered it offshore.

An abrupt, fourfold drop in ^{10}Be concentration occurs across the mid-Pleistocene transition at 0.8 Myr ago (Fig. 4c), a time when the duration and magnitude of glaciations increased²⁰. This drop might reflect reduced interglacial exposure or increasing erosivity of the ice sheet. ^{10}Be concentrations over the past 0.8 Myr are similar to those in sediments issuing from the western, southern and eastern ice margin of Greenland today^{18,21} (Figs 3 and 4c), except for one brief increase in a sample 400–200 kyr old. This higher concentration of ^{10}Be could reflect erosion of sediment exposed during the preceding long-lasting interglacial at MIS 11, a hypothesis supported by a slight, coincident rise in the $^{26}\text{Al}/^{10}\text{Be}$ ratio, consistent with interglacial re-exposure (Fig. 4c, d). With this exception, the consistently low ^{10}Be concentrations of the mid- to late Pleistocene sediment indicate the existence of a large, stable ice sheet in East Greenland for most of the past million years.

The shorter ^{10}Be record at Site 987 (2.2 Myr) is consistent with the latter part of the Site 918 record. Decay-corrected concentrations are steady and low at Site 987 (Fig. 4c, Supplementary Table 1), which suggests that IRD there was derived from glacial erosion of material that was deep below the land surface before East Greenland was ice-covered (Fig. 2a). Such efficient erosion of the former Miocene land surface is consistent with the location of Site 987 at the outlet of Scoresby Sund, a large fiord complex that is the major outlet for glaciers sourced in the highlands of East Greenland. Glacial erosion probably began early here because this part of East Greenland is thought to be where the ice sheet nucleated in the Late Miocene⁶ and where it survived even the most extreme Pliocene warm periods²². That the ^{10}Be concentrations at Site 918 are fourfold higher during the early Pleistocene than at Site 987 suggests that at least some of the sediment delivered to Site 918 over this interval was sourced from southeast Greenland.

Decay-corrected $^{26}\text{Al}/^{10}\text{Be}$ data from Site 918 provide additional information about the history of Greenlandic sediment (Fig. 4d). The lack of correlation between ^{10}Be concentration and $^{26}\text{Al}/^{10}\text{Be}$ (Extended Data Fig. 4f) suggests that changes in the $^{26}\text{Al}/^{10}\text{Be}$ ratio are not driven by long periods of surface exposure, because such exposure would raise both the ratio and ^{10}Be concentration. This lack of correlation is, however, consistent with changing sediment source areas and erosion of material with different $^{26}\text{Al}/^{10}\text{Be}$ ratios. Changes in sediment source area may be driven by changes in the basal thermal regime of the ice sheet instead of, or in addition to, changes in ice extent. Spatial and temporal differences in ice-sheet behaviour probably drive where and when subglacial erosion occurs because remote sensing data show that areas of warm (erosive) and cold (non-erosive) ice are closely juxtaposed²³.

$^{26}\text{Al}/^{10}\text{Be}$ does not decline steadily over time as would be expected if the entire source area of sediment had been completely and continually covered by ice, which would cause preferential loss of shorter-lived ^{26}Al (Fig. 4d). Rather, between 2.6 Myr and 1.7 Myr ago, the decay-corrected $^{26}\text{Al}/^{10}\text{Be}$ ratio (approximately 7.3) is mostly consistent with surface exposure. Decay-corrected $^{26}\text{Al}/^{10}\text{Be}$ ratios, similar to the production ratio, imply that ice did not cover the sediment source area for most of each glacial/interglacial cycle, because $^{26}\text{Al}/^{10}\text{Be}$ ratios change substantially only when surfaces are buried for several times longer than they are exposed and for at least several hundred thousand years in total²⁴. In contrast, most of the Site 918 record that is younger than 1.4 Myr

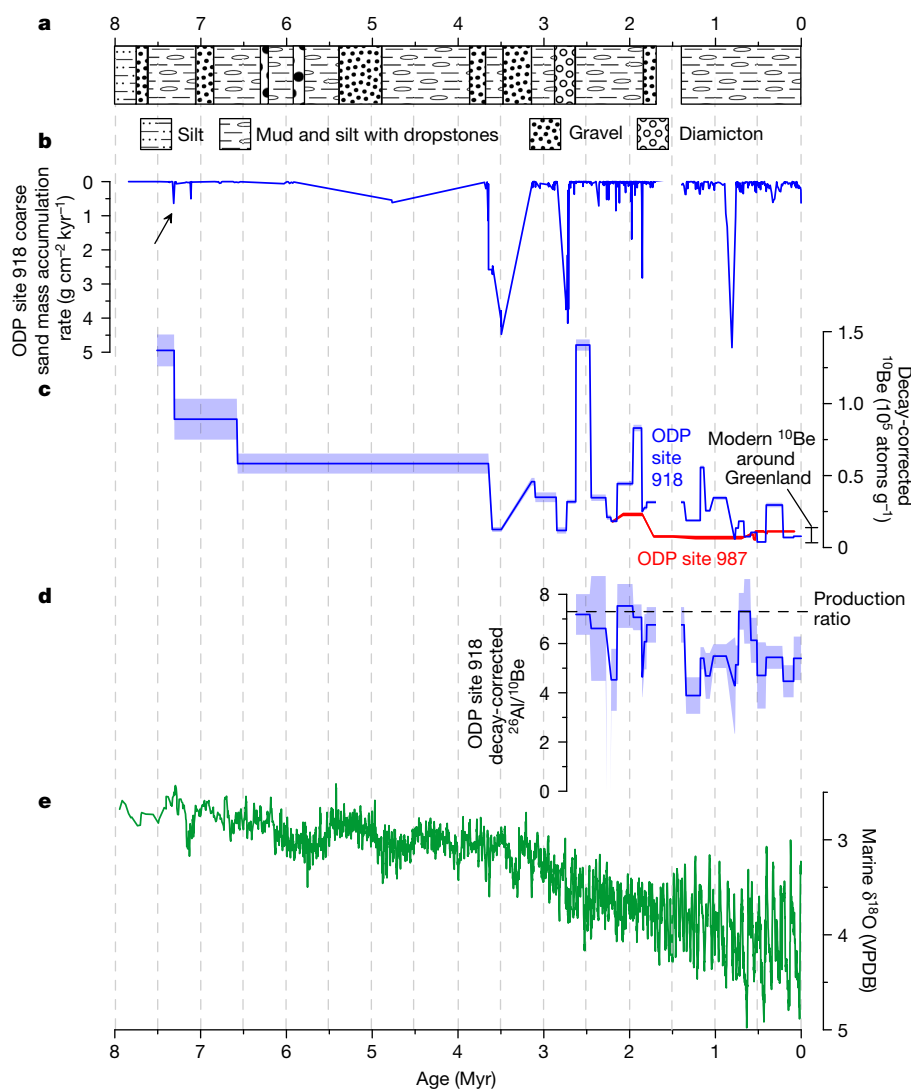


Figure 4 | Seven and a half million years of sediment cosmogenic-nuclide values from offshore East Greenland. **a**, Simplified lithostratigraphy at Site 918 (ref. 19). **b**, Coarse sand (0.25–2 mm) mass accumulation rate at Site 918 (ref. 19). Arrow indicates the oldest dropstones at the site⁶. **c**, Decay-corrected ^{10}Be concentrations at Sites

918 (blue, $n = 30$) and 987 (red, $n = 16$) with 1σ uncertainties. Black error bar shows the 1σ range of ^{10}Be in modern sediment ($n = 54$) from the Greenlandic margin¹⁸. **d**, Decay-corrected $^{26}\text{Al}/^{10}\text{Be}$ at Site 918 with 1σ uncertainty. **e**, Global marine $\delta^{18}\text{O}$ record²⁹. Note hiatus arising from missing core section between 1.7 Myr and 1.4 Myr (**a–d**).

has decay-corrected $^{26}\text{Al}/^{10}\text{Be}$ ratios of around 5 with no decrease over time, which is consistent with the excavation of sediment mostly buried under ice. However, at least some of the material eroded from 1.4 Myr ago to the present must have been intermittently exposed during interglacials in order to prevent the $^{26}\text{Al}/^{10}\text{Be}$ ratio from falling steadily because of radiodecay. For a short period just after the mid-Pleistocene transition and the largest IRD spike (samples 918-6 and 918-7), the decay-corrected $^{26}\text{Al}/^{10}\text{Be}$ rises to about 7.3; we infer that the change in climate cyclicity allowed ice in East Greenland to expand into and erode areas that were not recently ice-covered.

The best hope for detecting short periods of deglaciation is the $^{26}\text{Al}/^{10}\text{Be}$ ratio. Contemporary Greenlandic river sand, both glacially and non-glacially sourced, has a $^{26}\text{Al}/^{10}\text{Be}$ ratio of 7.6 ± 2.1 (1σ , $n = 5$), which is probably the result of landscape re-exposure during substantial mid-Holocene retreat²⁵ (Supplementary Table 2). Sand deposited in the Keglen delta at Kangerlussuaq during the end of the last glaciation around 7 kyr ago²⁶ has a $^{26}\text{Al}/^{10}\text{Be}$ ratio of 4.54 ± 0.58 , which is lower than the production ratio (Supplementary Table 2), and fully consistent with ratios we measured in marine cores over the last million years. Thus, high-precision $^{26}\text{Al}/^{10}\text{Be}$ ratio measurements of quartz extracted from a well dated, high-deposition-rate core may

reveal glacial–interglacial cycles and could be used to better assess the lag time between exposure and marine deposition²⁷.

Cosmogenic isotopes preserved in marine sediment record progressive erosion of the pre-glacial landscape in East Greenland from about 7.5 Myr to 2.7 Myr ago, the first growth of a full ice sheet at about 2.5 Myr, and a change in ice-sheet behaviour at the 0.8-Myr-ago mid-Pleistocene transition. Measuring ^{26}Al with ^{10}Be demonstrates that erosion under the East GIS varied over time and space, and suggests that during the early and mid-Pleistocene, the ice sheet in East Greenland expanded into previously ice-free terrain. Considered along with isotopic measurements of contemporary Greenlandic sediment¹⁸, the lack of repeated increases in ^{10}Be concentration or $^{26}\text{Al}/^{10}\text{Be}$ associated with interglacial periods during the past million years suggests that warming was sufficiently short-lived or modest that it seldom caused substantial and lengthy reductions in East GIS extent.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 February 2014; accepted 6 October 2016.

1. Nielsen, T. & Kuijpers, A. Only 5 southern Greenland shelf edge glaciations since the early Pliocene. *Sci. Rep.* **3**, 1875 (2013).
2. De Schepper, S., Gibbard, P. L., Salzmann, U. & Ehlers, J. A global synthesis of the marine and terrestrial evidence for glaciation during the Pliocene Epoch. *Earth Sci. Rev.* **135**, 83–102 (2014).
3. Gibbons, A. B., Megeath, J. D. & Pierce, K. L. Probability of moraine survival in a succession of glacial advances. *Geology* **12**, 327–330 (1984).
4. Funder, S. *et al.* Late Pliocene Greenland—the Kap København Formation in north Greenland. *Bull. Geol. Soc. Den.* **48**, 117–134 (2001).
5. Butt, A., Elverhøj, A., Forsberg, C. & Solheim, A. Evolution of the Scoresby Sund Fan, central East Greenland—evidence from ODP Site 987. *Norsk Geol. Tidsskr.* **81**, 3–15 (2001).
6. Larsen, H. C. *et al.* Seven million years of glaciation in Greenland. *Science* **264**, 952–955 (1994).
7. Helland, P. E. & Holmes, M. A. Surface textural analysis of quartz sand grains from ODP site 918 off the southeast coast of Greenland suggests glaciation of southern Greenland at 11 Ma. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **135**, 109–121 (1997).
8. Flesche Kleiven, H., Jansen, E., Fronval, T. & Smith, T. M. Intensification of Northern Hemisphere glaciations in the circum Atlantic region (3.5–2.4 Ma)—ice-rafted detritus evidence. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **184**, 213–223 (2002).
9. Bierman, P. R. *et al.* Preservation of a preglacial landscape under the center of the Greenland Ice Sheet. *Science* **344**, 402–405 (2014).
10. Yau, A. M., Bender, M. L., Blunier, T. & Jouzel, J. Setting a chronology for the basal ice at Dye-3 and GRIP: implications for the long-term stability of the Greenland Ice Sheet. *Earth Planet. Sci. Lett.* **451**, 1–9 (2016).
11. Lal, D. Cosmic ray labeling of erosion surfaces; in situ nuclide production rates and erosion models. *Earth Planet. Sci. Lett.* **104**, 424–439 (1991).
12. Granger, D. E. A review of burial dating methods using ^{26}Al and ^{10}Be . *Spec. Pap. Geol. Soc. Am.* **415**, 1–16 (2006).
13. Heisinger, B. *et al.* Production of selected cosmogenic radionuclides by muons. *Geochim. Cosmochim. Acta* **66**, A558 (2002).
14. Elverhøj, A., Hooke, R. L. & Solheim, A. Late Cenozoic erosion and sediment yield from the Svalbard–Barents sea region: implications for understanding erosion of glacierized basins. *Quat. Sci. Rev.* **17**, 209–241 (1998).
15. Bell, R. E. *et al.* Deformation, warming and softening of Greenland's ice by refreezing meltwater. *Nat. Geosci.* **7**, 497–502 (2014).
16. Sugden, D. E. & Watts, S. H. Tors, felsenmeer, and glaciation in northern Cumberland Peninsula, Baffin Island. *Can. J. Earth Sci.* **14**, 2817–2823 (1977).
17. DeConto, R. M. *et al.* Thresholds for Cenozoic bipolar glaciation. *Nature* **455**, 652–656 (2008).
18. Nelson, A. H., Bierman, P. R., Shakun, J. D. & Rood, D. H. Using in situ cosmogenic ^{10}Be to identify the source of sediment leaving Greenland. *Earth Surf. Process. Landf.* **39**, 1087–1100 (2014).
19. John, K. E. K. S. & Krissek, L. A. The late Miocene to Pleistocene ice-rafting history of southeast Greenland. *Boreas* **31**, 28–35 (2002).
20. Lisiecki, L. & Raymo, M. A Pliocene–Pleistocene stack of 57 globally distributed benthic $\delta^{18}\text{O}$ records. *Paleoceanography* **20**, PA1003–PA1020 (2005).
21. Goehring, B. M., Kelly, M. A., Schaefer, J. M., Finkel, R. C. & Lowell, T. V. Dating of raised marine and lacustrine deposits in east Greenland using beryllium-10 depth profiles and implications for estimates of subglacial erosion. *J. Quat. Sci.* **25**, 1–10 (2010).
22. Hill, D. J., Dolan, A. M., Haywood, A. M., Hunter, S. J. & Stoll, D. K. Sensitivity of the Greenland Ice Sheet to Pliocene sea surface temperatures. *Stratigraphy* **7**, 111–122 (2010).
23. Petrunin, A. G. *et al.* Heat flux variations beneath central Greenland's ice due to anomalously thin lithosphere. *Nat. Geosci.* **6**, 746–750 (2013).
24. Bierman, P. R., Davis, P. T., Corbett, L. B., Lifton, N. & Finkel, R. Cold-based, Laurentide ice covered New England's highest summits during the Last Glacial Maximum. *Geology* **43**, 1059–1062 (2015).
25. Larsen, N. K. *et al.* The response of the southern Greenland ice sheet to the Holocene thermal maximum. *Geology* **43**, 291–294 (2015).
26. Storms, J. E. A., de Winter, I. L., Overeem, I., Drikkoningen, G. G. & Lykke-Andersen, H. The Holocene sedimentary history of the Kangerlussuaq Fjord-valley fill, West Greenland. *Quat. Sci. Rev.* **35**, 29–50 (2012).
27. DePaolo, D. J., Maher, K., Christensen, J. N. & McManus, J. Sediment transport time measured with U-series isotopes: results from ODP North Atlantic drift site 984. *Earth Planet. Sci. Lett.* **248**, 394–410 (2006).
28. Balco, G., Stone, J., Lifton, N. & Dunai, T. A complete and easily accessible means of calculating surface exposure ages or erosion rates from ^{10}Be and ^{26}Al measurements. *Quat. Geochronol.* **3**, 174–195 (2008).
29. Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements Research supported by NSF ARC-1023191. A. Nelson prepared some samples. W. Hale and the Bremen Core Repository facilitated core sampling. G. Balco provided input on muon production. We thank K. St John for providing ODP site 918 mass accumulation rate data, B. de Boer for ice sheet model output, W. Huang for running foraminifer stable isotope samples, and S. Xu and the staff of the SUERC AMS laboratory for support during ^{26}Al measurements. This is LLNL-JRNL-701099.

Author Contributions P.R.B. and J.D.S. designed the experiment. J.D.S. oversaw core sampling. P.R.B. and L.B.C. did and oversaw laboratory work. D.H.R., S.R.Z. and P.R.B. performed isotopic analyses. P.R.B., J.D.S., L.B.C. and D.H.R. interpreted the data and all authors contributed to the preparation of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.R.B. (pbierman@uvm.edu).

Reviewer Information *Nature* thanks D. Dahl-Jensen, D. Granger and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Hypothesis testing. We use new isotopic data, in conjunction with sensitivity tests, a forward model, and other extant records, to evaluate three hypotheses about the behaviour of the East GIS that previous data have not been able to address conclusively. For the past 7.5 Myr, we test whether (1) East GIS behaviour mirrored global climate/ice volume as represented by the marine $\delta^{18}\text{O}$ record; (2) the efficacy of erosion under the East GIS varied over time and space; (3) most interglacial periods were sufficiently short-lived or cool enough that they did not cause notable reductions in East GIS extent.

Compilation of deep time history of the GIS. Most of what is known about long-term ice-sheet history comes from marine sediment records interpreted as global or regional proxies for ice volume or glacial activity. For example, stable oxygen isotope measurements of foraminifera isolated from marine sediment track global ice volume and ocean temperature, but provide little information about the individual behaviour of each of the world's major ice sheets²⁰. Global sea-level history reflects total ice volume, but in a complex fashion³⁰ because the record is aliased by local tectonic and glacio-isostatic adjustment of land levels³¹. The most robust inferences about the comings and goings of now-vanished ice sheets are based on the presence and provenance in marine sediment of IRD shed from melting icebergs that originated on glaciated continents^{2,8}. IRD records are illustrative of when sediment-bearing glacial ice reached the coast, but with few exceptions³² do not otherwise constrain ice extent³³. Figure 1 presents our compilation of references relevant to understanding the history of ice on Greenland since the Miocene. References for Fig. 1 are cited in Extended Data Fig. 5.

Determining sediment source area. Making accurate inferences about ice-sheet behaviour on the basis of terrestrial sediment recovered from marine archives requires knowledge of the sediment source area. Multiple lines of evidence indicate that the quartz we isolated was sourced from East Greenland. The East Greenland Current (Fig. 3) drifts icebergs from north to south over both Sites 918 and 987, which suggests that the IRD we analysed is dominantly from East Greenland³⁴. IRD composition downcore at Site 918 consistently indicates eastern Greenland sediment sourcing for millions of years^{35–37}. While there may be some contribution from gravity flows off the continental shelf, sedimentological evidence suggests that most sand at Site 918 comes from ice rafting rather than turbidites^{6,19,38,39}. At Site 918, sand is compositionally similar to larger dropstones, which is consistent with an IRD source for the sand⁴⁰. Sediment at Site 987 is probably more locally sourced because drilling was done on the toe of a large subaqueous fan⁵; although some of the Site 987 sediment may come from the north, most was presumably delivered directly from ice flowing east through Scoresby Sund (Fig. 3). In summary, the cosmogenic data we present reflect the history of and processes active in eastern Greenland.

Thermal conditions at the base of the ice sheet are not well known and change over time⁴¹ and space^{23,42}. Warm-based ice (the ultimate source of the sediment we analysed because it is required to erode the material) is most likely to be found in deep troughs, near some ice margins, and where geothermal heat flux is high^{23,43,44}. Models suggest that 20% to 30% of the pre-industrial Holocene GIS was warm-based⁴³, but during the Last Glacial Maximum up to 50% of ice on Greenland may have been warm-based, perhaps due to increased thickness⁴¹.

Paired $^{26}\text{Al}/^{10}\text{Be}$ approach and the $^{26}\text{Al}/^{10}\text{Be}$ production ratio of 7.3. Because cosmogenic nuclides with different half-lives decay at different rates after production ceases, multiple nuclides can be measured in tandem (for example, ^{10}Be and ^{26}Al) to provide insight about periods of burial. A multi-nuclide approach can thus constrain the timing and duration of burial by non-erosive, cold-based ice⁴⁵, which is a process that has probably occurred variably in Greenland over both space and time.

When exposure begins on a fresh surface, the $^{26}\text{Al}/^{10}\text{Be}$ ratio is the production ratio of the two nuclides. If a previously exposed surface is buried and shielded from further nuclide production, the $^{26}\text{Al}/^{10}\text{Be}$ ratio drops because the 0.71-Myr half-life of ^{26}Al (ref. 46) is shorter than the 1.39-Myr half-life of ^{10}Be (refs 47–49). If a sample is exposed again following burial, production resumes and the $^{26}\text{Al}/^{10}\text{Be}$ ratio increases because the production rate of ^{26}Al is greater than that of ^{10}Be . It is important to note that relatively short burial durations (<100–200 kyr) and/or re-exposure following burial can result in $^{26}\text{Al}/^{10}\text{Be}$ ratios that are indistinguishable from the production ratio^{24,50} even though the surface has experienced periods of burial lasting tens of thousands of years.

Any inferences stemming from $^{26}\text{Al}/^{10}\text{Be}$ ratios are largely dependent upon the assumed $^{26}\text{Al}/^{10}\text{Be}$ production ratio, which is a direct function of the production rates of the two nuclides. Although nuclide production rates have long been known to vary across latitude and elevation^{11,51}, it has generally been assumed that ^{10}Be and ^{26}Al production rates scale similarly, with a resulting production ratio of 6.75 for all locations on Earth's surface²⁸. However, recent work has suggested that the production ratio is itself dependent on latitude and elevation because each isotope's production rate scales differently across space^{52–54}. Argento *et al.*⁵² used numerical

models to estimate a $^{26}\text{Al}/^{10}\text{Be}$ production ratio of 7.0 to 7.1 at sea level and high latitude, which is in agreement with the median value of 7.16 calculated from low-elevation (<2,000 m) calibration samples presented in the same study. Sites from a range of latitudes and elevations have production ratios ranging from 7.0 to 7.3, scaled to sea level and high latitude, and using seven different scaling schemes⁵⁴. Atmospheric mass drives the differences in production between nuclides, with elevation probably being more important than latitude⁵⁵, although comprehensive studies of the global variation in the $^{26}\text{Al}/^{10}\text{Be}$ production ratio have not yet been conducted.

In this study, we place more emphasis on the relative rather than the absolute $^{26}\text{Al}/^{10}\text{Be}$ ratio in marine sediment over time; hence the assumed $^{26}\text{Al}/^{10}\text{Be}$ production ratio is less important here than in studies inferring absolute exposure and burial durations. However, we base our assumed production ratio on the work of Corbett *et al.*⁵⁶, who quantified $^{26}\text{Al}/^{10}\text{Be}$ in 24 continuously exposed bedrock and boulder surfaces at four high-latitude sites in Greenland that were deeply eroded during the last glaciation. They determined a $^{26}\text{Al}/^{10}\text{Be}$ production ratio of 7.3 ± 0.3 (slope of a York linear regression fit to all data with errors in both variables, 1σ), supporting recent modelling work indicating that the production ratio can exceed 6.75 at locations such as Greenland. Although the geographic variability of the production ratio is still unclear, we choose to employ the production ratio of ref. 56 here because the source of the sediments from Sites 918 and 987 is similar to the latitude range of the calibration samples in their data set.

Sample measurements. We measured ^{10}Be in 30 samples and ^{26}Al in 22 samples spanning the last 7.5 Myr and 2.6 Myr, respectively, in sediment cores at Site 918, located in the Irminger basin 110 km southeast of Greenland (63.1°N , 38.6°W , 1,800 m water depth). This site was previously used to define the onset of Greenland glaciation on the basis of the earliest occurrence of IRD⁶, which is included in our oldest sample. We also measured ^{10}Be in 16 samples from Site 987 spanning the last 2.2 Myr of deposition 130 km offshore of Scoresby Sund and 1,200 km northeast of Site 918 (70.5°N , 17.9°W , 1,670 m water depth)⁵⁷.

Core samples were obtained from the Bremen Core Repository. We disaggregated and wet-sieved sediments isolating the grain size fraction 0.125–0.750 mm and used weak acid ultrasonic leaching (0.25% to 0.5% HF and HNO_3) to slowly dissolve all minerals other than quartz⁵⁸. We amalgamated quartz from subsamples taken over an interval of core until we had sufficient quartz mass (7.8 g to 25.3 g) from which to extract and reliably measure ^{10}Be . Thus, samples represent the average ^{10}Be content of quartz present in core sections ranging in length from 0.04 m to 91 m (median 6 m, standard deviation 19 m). All uncertainties reported in this paper are 1σ .

Age spans for samples range from 0.001 Myr to 2.9 Myr (median 0.1 Myr, standard deviation 0.5 Myr). Our marine sediment record of ^{10}Be and ^{26}Al concentrations does not have the temporal resolution to clearly reflect major high-frequency changes in Plio-Pleistocene climate, such as the major interglacials at MIS 11, MIS 9 or MIS 5e. The need to amalgamate sufficient quartz for measuring very low isotope abundances meant that integration of core sediment over depth (and thus time) mixed sand deposited during glacial and interglacial periods; analysis of a core more proximal to the continental shelf might overcome this limitation.

After purifying quartz, samples were dissolved using HF in the presence of ^9Be carrier produced from beryl. Samples were processed in batches of 12 including two full chemistry process blanks⁵⁹. ^{10}Be measurements were made at the Center for Accelerator Mass Spectrometry at Lawrence Livermore National Laboratory^{60,61} and referenced to standard 07KNSD3110 (ref. 46) assuming a $^{10}\text{Be}/^9\text{Be}$ ratio of $2,850 \times 10^{-15}$. ^{26}Al measurements were made at the Scottish Universities Environmental Research Centre⁶² and normalized to the Z92-0222⁴⁶ standard with a nominal $^{26}\text{Al}/^{27}\text{Al}$ ratio of 4.11×10^{-11} . The average blank ratio ($^{10}\text{Be}/^9\text{Be} = (4.6 \pm 1.0) \times 10^{-16}$, $n = 6$; group 1, $^{26}\text{Al}/^{27}\text{Al} = (8.5 \pm 2.1) \times 10^{-16}$, $n = 4$; group 2 $^{26}\text{Al}/^{27}\text{Al} = (14.9 \pm 4.5) \times 10^{-16}$, $n = 4$) was subtracted from measured ratios, and uncertainties in sample and blank ratios were propagated in quadrature (Supplementary Table 1).

Replicate preparation of sample 918-17 (918-17X) indicates reproducibility within measurement uncertainty (Supplementary Table 1). Statistically identical measured concentrations of ^{10}Be in four samples (987-E to 987-H) collected from different depths in a 70-cm-thick IRD-rich layer ($4,250 \pm 370$ to $4,460 \pm 300$ atoms g^{-1}) also demonstrate the reproducibility of our measurements (Supplementary Table 1).

In all samples, measured ^{10}Be concentrations are low (2,100 to 40,000 atoms g^{-1}), but well above procedural backgrounds. Because of the shorter half-life of ^{26}Al , it is detectable only in younger samples (<2.6 Myr), and was measured only at Site 918; concentrations of ^{26}Al are also low (9,700 to 118,000 atoms g^{-1} ; Supplementary Table 1), but similarly well above background. Cosmogenic $^{26}\text{Al}/^{10}\text{Be}$ ratios at the time of deposition (corrected by core depth–age models) range from about 3.9 to about 7.5 (Supplementary Table 1).

Age models for Sites 918 and 987 and decay-correction procedure. For Site 918, we used established age–depth constraints from ref. 19, who applied ages from the timescale of ref. 63 to magnetostratigraphic⁶⁴ and biostratigraphic datums^{65,66}. Ages were linearly interpolated between these control points (Extended Data Fig. 1). Note that there is an erosional hiatus at 71.1 m below the seafloor, which is estimated to span 1.71–1.39 Myr¹⁹. We also developed a planktonic $\delta^{18}\text{O}$ record (*N. pachyderma*, left-coiling) to refine the age model above the Brunhes–Matuyama reversal (780 kyr) at 45.9 m below the seafloor⁶⁴. 168 stable isotope measurements were made at the Lamont–Doherty Earth Observatory, and 11–15 tests were used per sample. The $\delta^{18}\text{O}$ record clearly displays the Holocene and the last interglacial, but there is some ambiguity in the identification of other marine isotope stages, such as MIS 11 and MIS 13 (Extended Data Fig. 2).

For Site 987, we developed an age model by linearly interpolating between the age control points reported by the Leg 162 shipboard scientific party⁵⁷, which are primarily based on palaeomagnetic events (Extended Data Fig. 1).

Measured ^{10}Be and ^{26}Al concentrations (Supplementary Table 1) were corrected for decay since deposition on the seafloor using these age models and assuming half-lives of ^{10}Be $t_{1/2} = 1.39$ Myr (ref. 49) and ^{26}Al $t_{1/2} = 0.71$ Myr (ref. 67). Since our cosmogenic-nuclide samples were amalgamated from subsamples spanning 0.001 Myr to 2.9 Myr (Supplementary Table 3), we used the sand mass-weighted mean age of these subsamples to derive a single integrated age for each cosmogenic sample. Age model uncertainties can alter the absolute value of decay-corrected ^{10}Be concentrations and change the timing of some isotopic shifts, but have minimal impact on the overall structure of the record.

Sensitivity tests. We examined the sensitivity of ^{10}Be concentrations and $^{26}\text{Al}/^{10}\text{Be}$ ratios to erosion, burial, exposure and mixing (Fig. 2), assuming sea-level high-latitude production rates, including production from muons calculated using the MATLAB implementation^{13,28}. Depth profiles were first run to secular equilibrium, which was reached when nuclide production balanced loss via radiodecay and erosion; the latter was simulated by shifting the profile upward each time step in proportion to the prescribed ice-free surface erosion rate (5 m Myr^{−1}, 20 m Myr^{−1} or 50 m Myr^{−1}). Steady-state profiles with higher erosion rates have lower ^{10}Be concentrations because nuclides are shed more rapidly, but they have higher $^{26}\text{Al}/^{10}\text{Be}$ ratios because nuclides are brought to the surface more quickly and thus have less time to decay in the subsurface (Fig. 2a). We simulated cold-based ice cover for 1 Myr by halting production and allowing the 20 m Myr^{−1} steady-state profile to decay in place without additional sub-ice erosion, whereas an analogous simulation for warm-based ice cover continued to erode under ice at 20 m Myr^{−1}. Surface nuclide concentrations decrease much faster under the erosive warm-based ice, and $^{26}\text{Al}/^{10}\text{Be}$ ratios also decline more quickly since the erosive ice brings deeper-buried (and thus longer-buried) nuclides to the surface. The 1-Myr-long warm-based ice simulation was performed again, but interrupted by either a 10-kyr or 200-kyr episode of interglacial exposure (with ice-free surface erosion continuing at 20 m Myr^{−1}) halfway through the simulation. In these simulations, because nuclide concentrations were very low before the interglacials, both were able to quickly reset the $^{26}\text{Al}/^{10}\text{Be}$ ratio to preglacial values; however, only the very long (200 kyr) interglacial had sufficient time to fully rebuild nuclide concentrations. Lastly, we modelled the mixing of sediment from low-concentration, low-ratio (eroded and long-buried) and high-concentration, high-ratio (long-exposed) endmembers to understand how the values we measured in marine sediments might reflect contributions from multiple source areas on Greenland. Nuclide concentrations mix linearly: $C_{\text{mixed}} = C_1F_1 + C_2F_2$, where C_1 and C_2 and F_1 and F_2 are the nuclide concentrations and mixing fractions ($F_1 + F_2 = 1$) of the two endmembers. $^{26}\text{Al}/^{10}\text{Be}$ ratios, however, exhibit nonlinear mixing that is weighted by the ratio of the endmembers' nuclide concentrations, because the greater the number of nuclides one endmember contributes relative to the other, the more it influences the mixed nuclide ratio: $R_{\text{mixed}} = R_1(C_1F_1/(C_1F_1 + C_2F_2)) + R_2(C_2F_2/(C_1F_1 + C_2F_2))$.

Our sensitivity tests demonstrate how progressively deeper erosion, interglacial exposure, burial by cold-based ice, and sediment mixing from different sources affect the concentration of ^{10}Be and ^{26}Al in terrestrial sediment exported from Greenland (Fig. 2). Such modelling shows that covering a landscape with non-erosive, cold-based ice for hundreds of thousands of years lowers the $^{26}\text{Al}/^{10}\text{Be}$ ratio, but does not much change ^{10}Be concentration because of the long half-life of ^{10}Be in relation to the burial duration (Fig. 2b). In contrast, cover by erosive, warm-based ice not only lowers the $^{26}\text{Al}/^{10}\text{Be}$ ratio by shielding the bed from cosmic ray exposure, but also lowers nuclide concentrations because it erodes material with previously produced nuclides and incorporates rock or sediment that was once deeply shielded from cosmic radiation. After the upper several metres of rock and soil are eroded by warm-based ice, isotopic concentrations in the resulting sediment are low and relatively insensitive to continued erosion. This is because the concentration of ^{10}Be in sediment produced by

glaciers is controlled primarily by the extent of sub-ice erosion into the deep, muon-dominated production zone that extends tens of metres below the preglacial land surface where nuclide concentration changes only gradually with depth (Fig. 2a). When sediment is the result of mixing of components with different burial and erosion histories, the history of the sediment may be constrained by considering possible end members with different nuclide concentrations and $^{26}\text{Al}/^{10}\text{Be}$ ratios, mixed in different proportions (Fig. 2d). The ^{10}Be concentrations we measured reflect the erosion-weighted average ^{10}Be concentration of the areas from which they were sourced, while $^{26}\text{Al}/^{10}\text{Be}$ ratios are biased towards source areas that had relatively high nuclide concentrations.

^{10}Be and ^{26}Al concentrations measured in contemporary terrestrial Greenlandic sediment. To better constrain the interpretation of cosmogenic-nuclide measurements in marine sediment, we collected sediment samples from Greenlandic rivers, moraines and river terraces and measured their ^{10}Be (ref. 18), and in some cases, ^{26}Al concentrations (Supplementary Table 2). Sediment sourced from the ice sheet in eastern, western and southern Greenland both today¹⁸ and at the end of the last glaciation (sampled in well-dated terraces)^{18,21} has very low concentrations of ^{10}Be of only thousands of atoms per gram. Sediment in streams draining only areas outside the current ice margin has on average several times more ^{10}Be , which reflects exposure of the land surface to cosmic radiation during the Holocene¹⁸. Isotope and mass balance calculations indicate that most sediment now being delivered to the Greenlandic margin originates from beneath the ice sheet and not from the deglaciated margin¹⁸.

To complement existing ^{10}Be data¹⁸, we measured ^{26}Al in four samples of contemporary river sediment as well as sediment from the Keglen Delta terrace at Kangerlussuaq²⁶ (sample GLX-08) and another terrace deposited near Narsarsuaq⁶⁸ (GLX-34). Sediment in the Keglen Delta was deposited during the deglaciation (about 7 kyr ago)²⁶ and has a $^{26}\text{Al}/^{10}\text{Be}$ ratio substantially lower than the production ratio (GLX-08, 4.54 ± 0.58 , 1σ). All sediment from modern streams, as well as that in the terrace at Narsarsuaq (GLX-34) deposited during a neoglacial readvance about 1.5 kyr (after mid Holocene retreat and exposure of the now mostly ice-covered landscape)⁶⁸, has an average ratio of 7.62 ± 2.12 (1σ ; $n = 5$), similar to the production ratio. These data imply that at deglaciation, sediment leaving the ice sheet about 7 kyr ago had an $^{26}\text{Al}/^{10}\text{Be}$ ratio lower than the production value, and that exposure during the mid-Holocene, when the GIS retreated a few to tens of kilometres inland of the current margin, raised the $^{26}\text{Al}/^{10}\text{Be}$ to or near that of production, as suggested by our sensitivity tests (Fig. 2).

These results imply that short periods (about 10–20 kyr) of subaerial interglacial exposure, primarily at the margins of the ice sheet, matter little because they change the nuclide concentration substantially only in the uppermost few metres of rock or soil via shallow neutron spallation reactions. However, even short interglacial re-exposure can effectively raise the $^{26}\text{Al}/^{10}\text{Be}$ ratio if initial nuclide concentrations are very low when exposure begins (Fig. 2c).

Inherent method limitations. The glacial sediment system itself may limit the resolution of the record. Sediment tracing using ^{10}Be unambiguously shows that most sediment delivered to the current-day Greenland margin during the Holocene interglacial is derived from under the ice, has very little ^{10}Be and ^{26}Al , and is not sourced from the deglaciated peripheral area¹⁸. Sediment currently being shed from deglaciated terrain has concentrations of ^{10}Be several times higher than that of glacially derived material, but the marginal area is small in comparison to the area still covered by ice¹⁸. During glacial advances, sediment from previously exposed margins will be incorporated by ice and eventually mixed with long-shielded material and moved offshore. Even though the marginally sourced material has higher concentrations of ^{10}Be , it is overwhelmed volumetrically by material coming from areas that have been long covered by ice and thus limits the marine record's sensitivity to interglacial cosmic-ray exposure.

Calculation of background, pre-glaciation erosion rate for East Greenland. We determined a pre-glacial erosion rate for southeast Greenland from the decay-corrected ^{10}Be concentration in our oldest sample at Site 918 ($135,000 \pm 10,900$, 1σ ; Supplementary Table 1), which integrates sediments from the 20 m of core immediately below the oldest dropstone at Site 918 identified by ref. 6. Assuming the ^{10}Be in this sample was produced at the surface at sea level directly onshore from Site 918 and experienced no topographic shielding, we obtain an erosion rate of approximately 22 ± 3 m Myr^{−1} using the CRONUS calculator v.2.3²⁸. This estimate is relatively insensitive to these assumptions, except for elevation, which would, for instance, double the erosion rate if production occurred at 1,000 m above sea level rather than at sea level.

Site 918 ^{10}Be comparison to core sand content. We measured Site 918 sand ($>63\mu\text{m}$) content and binned data over the same depth intervals as the ^{10}Be samples to facilitate comparison. We similarly binned values in the marine benthic $\delta^{18}\text{O}$ record²⁹ over the same time intervals as the Site 918 ^{10}Be record. Regressions, using logarithmic scaling for the Site 918 ^{10}Be and sand records, show pronounced

relationships, with lower ^{10}Be associated with higher sand content and more enriched marine $\delta^{18}\text{O}$ ($r^2 = 0.52$ in both cases, $P < 0.001$) (Extended Data Fig. 3). As Site 918 sand concentrations probably reflect glacial erosion on land and marine $\delta^{18}\text{O}$ is a proxy for global ice volume, these relationships are broadly consistent with intensified glacial activity yielding lower ^{10}Be concentrations in East-Greenland-derived sediments.

Forward modelling and code availability. As a first-order attempt to reproduce the Sites 918 and 987 cosmogenic-nuclide records, we constructed a simple model of GIS dynamics and cosmogenic-nuclide concentrations driven by three different plausible ice volume reconstructions over the past 5.3 Myr. The model consists of two sets of ten parallel cosmogenic-nuclide depth profiles, as described in Methods section 'Sensitivity tests', and was initialized using 20 m Myr^{-1} ice-free surface erosion rate steady-state nuclide depth profiles reflecting pre-ice-sheet conditions as indicated by the deepest sample in Site 918. Ice sheet extent was modelled from 0% to 100% in 10% increments by turning nuclide production on or off for the corresponding number of depth profiles at a given time step; for example, production was on for all depth profiles when ice cover was 0%, but for only nine profiles when ice cover was 10%, and so on. The time step is 2 kyr. Since the actual GIS extent through time is poorly constrained, we tried parameterizing it with three different time series: a sea-level record derived from $\delta^{18}\text{O}$ variations in the semi-enclosed Mediterranean Sea basin⁶⁹, the marine $\delta^{18}\text{O}$ record of global ice volume and deep ocean temperature²⁰, and a simulated history of the GIS from an ice-sheet model forced by the marine $\delta^{18}\text{O}$ record⁷⁰. The last time series explicitly gives ice-sheet extent; the relationship between the first two series and GIS extent was calibrated by assuming that ice cover was 100% at 12 kyr, 80% today, 50% during MIS 11, 20% during the mid-Pliocene, and 0% in the Miocene.

We used a simple formulation of basal temperature regimes beneath the modelled ice. Because the GIS has roughly equal areas of cold- and warm-based ice today⁴⁴, we set the modelled ice cover also to have equal fractions by making one set of depth profiles warm-based (sub-ice erosion rate 20 m Myr^{-1}) and the other set cold-based (sub-ice erosion rate 0 m Myr^{-1}). Spatial variability in basal temperature regimes was introduced by switching between the warm-based and cold-based regime of the two sets of depth profiles every 500 kyr; this is not meant to be realistic, but rather simply to help assess the role of this variable in driving cosmogenic-nuclide concentrations given that the basal thermal history of the GIS is not known. Erosion rates were 20 m Myr^{-1} in ice-free areas. The simulated cosmogenic-nuclide values shown in Extended Data Fig. 4 represent the material shed from ice-covered, warm-based depth profiles in the model, and assume instantaneous transport to the deep sea.

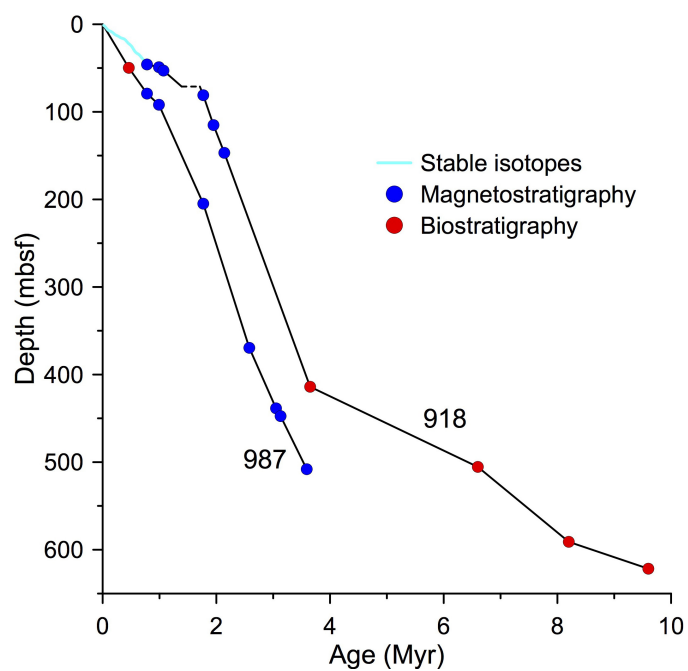
This forward modelling illustrates the limitations in the approach we present here as well as the uncertainty of assumptions underlying the model (Extended Data Fig. 4). Our model reproduces the overall ^{10}Be record for both Sites 918 and 987, but does not capture the fine structure of the Site 918 data. The ice-sheet extent from a model simulation⁷⁰ consistently underestimates ^{10}Be and $^{26}\text{Al}/^{10}\text{Be}$, probably because it does not accurately reflect GIS dynamism in the Pleistocene. The marine $\delta^{18}\text{O}$ proxy²⁰ and sea-level proxy⁶⁹ generate more realistic $^{26}\text{Al}/^{10}\text{Be}$; the sea-level proxy generates the best fit to the ^{10}Be record. We interpret the fine structure (^{10}Be peaks at 2.5 Myr, 1.9 Myr and 1.1 Myr), which we cannot model, as changes in the sediment source area to which cosmogenic nuclides are singularly sensitive; most likely these peaks represent expansion of warm-based areas of the ice sheet into terrain that had not previously been eroded.

The MATLAB code files used to generate the forward model are available at <https://github.com/shakun/Bierman-et-al-2016-Nature>. The three versions of the model are provided as MATLAB code files with the forcing series representing GIS extent through time designated in the model file name (*forward_model_XXXXX.mat*). These input driving series are the deep-sea $\delta^{18}\text{O}$ record²⁰ (*LR04.mat*), the Mediterranean Sea sea-level record⁶⁹ (*med.mat*), and simulated ice sheet extent based on modelling⁷⁰ (*deboer.mat*), all given at 2-kyr resolution over the past 5.3 Myr. Initialized bedrock profiles with steady-state ^{10}Be and ^{26}Al concentrations at 1-cm depth increments below the surface assuming a sea-level high-latitude production rate and 20 m Myr^{-1} ice-free surface erosion rate are given in *steadystate_10Be_20mMyr.mat* and *steadystate_26Al_20mMyr.mat*. Sea-level high-latitude ^{10}Be and ^{26}Al production rates in 1-cm depth increments below the surface are given in *P10.mat* and *P26.mat*. The file *er_half_Ma.mat* determines which set of bedrock profiles are beneath erosive warm-based ice (1) or non-erosive cold-based ice (0) at each time step.

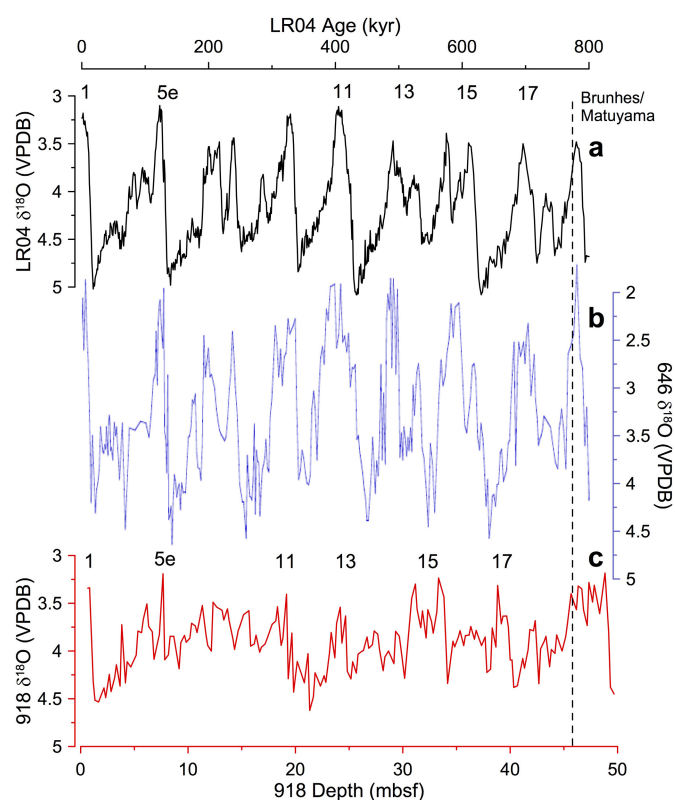
Data availability. All data generated and analysed during this study are included in this Letter and its Supplementary Information. MATLAB forward-model code is available at <https://github.com/shakun/Bierman-et-al-2016-Nature>.

30. Spratt, R. M. & Lisiecki, L. E. A Late Pleistocene sea level stack. *Clim. Past Discuss.* **11**, 3699–3728 (2015).
31. Clark, P. U. & Mix, A. C. Ice sheets and sea level of the last glacial maximum. *Quat. Sci. Rev.* **21**, 1–7 (2002).
32. Hemming, S. R., Bond, G. C., Broecker, W. S., Sharp, W. D. & Klas-Mendelson, M. Evidence from $^{40}\text{Ar}/^{39}\text{Ar}$ ages of individual hornblende grains for varying Laurentide sources of iceberg discharges 22,000 to 10,500 yr B.P. *Quat. Res.* **54**, 372–383 (2000).
33. Dowdeswell, J. A., Cofaigh, C. Ó., Andrews, J. T. & Scourse, J. D. Workshop explores debris transported by icebergs and paleoenvironmental implications. *Eos* **82**, 382–386 (2001).
34. Martin, T. & Wadhams, P. Sea-ice flux in the East Greenland Current. *Deep Sea Res. Part II* **46**, 1063–1082 (1999).
35. Bridgewater, D., Keto, L., McGregor, V. R. & Myers, J. S. in *Geology of Greenland* (eds Escher, E. & Watt, W. S.) 304–339 (Geological Survey of Greenland, 1976).
36. Larsen, H. C. Geological perspectives of the East Greenland continental margin. *Bull. Geol. Soc. Den.* **29**, 77–101 (1980).
37. Linthout, K., Troelstra, S. R. & Kuijpers, A. Provenance of coarse ice-rafted detritus near the SE Greenland margin. *Netherlands J. Geosci.* **79**, 109–121 (2000).
38. Spezzaferri, S. L. in *Proc. ODP Sci. Res. Vol. 152* (eds Larsen, H. C., Saunders, A. & Clift, P. D.) 161–190 (Ocean Drilling Program, 1998).
39. Molnia, B. F. in *Glacial-Marine Sedimentation* (ed. Molnia, B. F.) 593–626 (Plenum, 1983).
40. Party, S. S. in *Proc. ODP Init. Rep. Vol. 152* (eds Larsen, H. C., Saunders, A. & Clift, P. D.) 177–256 (Ocean Drilling Program, 1994).
41. Greve, R. Relation of measured basal temperatures and the spatial distribution of the geothermal heat flux for the Greenland ice sheet. *Ann. Glaciol.* **42**, 424–432 (2005).
42. Kaus, B. J. P. Heating glaciers from below. *Nat. Geosci.* **6**, 683–684 (2013).
43. Fyke, J. G., Sacks, W. J. & Lipscomb, W. H. A technique for generating consistent ice sheet initial conditions for coupled ice sheet/climate models. *Geosci. Model Dev.* **7**, 1183–1195 (2014).
44. MacGregor, J. A. et al. A synthesis of the basal thermal state of the Greenland Ice Sheet. *J. Geophys. Res.* **121**, 1328–1350 (2016).
45. Bierman, P., Marsella, K., Patterson, C., Davis, P. & Caffee, M. Mid-Pleistocene cosmogenic minimum-age limits for pre-Wisconsinan glacial surfaces in southwestern Minnesota and southern Baffin Island: a multiple nuclide approach. *Geomorphology* **27**, 25–39 (1999).
46. Nishiizumi, K. Preparation of ^{26}Al AMS standards. *Nucl. Instrum. Methods Phys. Res. B* **223/224**, 388–392 (2004).
47. Nishiizumi, K. et al. Absolute calibration of ^{10}Be AMS standards. *Nucl. Instrum. Methods Phys. Res. B* **258**, 403–413 (2007).
48. Chmieleff, J., Von Blanckenburg, F., Kossert, K. & Jakob, D. Determination of the ^{10}Be half-life by multicollector ICP-MS and liquid scintillation counting. *Nucl. Instrum. Methods Phys. Res. B* **268**, 192–199 (2010).
49. Korschinek, G. et al. A new value for the half-life of ^{10}Be by heavy-ion elastic recoil detection and liquid scintillation counting. *Nucl. Instrum. Methods Phys. Res. B* **268**, 187–191 (2010).
50. Corbett, L., Bierman, P. & Rood, D. Constraining multi-stage exposure-burial scenarios for boulders preserved beneath cold-based glacial ice in Thule, northwest Greenland. *Earth Planet. Sci. Lett.* **440**, 147–157 (2016).
51. Stone, J. Air pressure and cosmogenic isotope production. *J. Geophys. Res.* **105**, 23753–23759 (2000).
52. Argento, D., Reedy, R. & Stone, J. Modeling the earth's cosmic radiation. *Nucl. Instrum. Methods Phys. Res. B* **294**, 464–469 (2013).
53. Argento, D., Stone, J., Reedy, R. & O'Brien, K. Physics-based modeling of cosmogenic nuclides part II—key aspects of in-situ cosmogenic nuclide production. *Quat. Geochronol.* **26**, 44–55 (2015).
54. Borchers, B. et al. Geological calibration of spallation production rates in the CRONUS-Earth project. *Quat. Geochronol.* **31**, 188–198 (2016).
55. Lifton, N., Sato, T. & Dunai, T. J. Scaling in situ cosmogenic nuclide production rates using analytical approximations to atmospheric cosmic-ray fluxes. *Earth Planet. Sci. Lett.* **386**, 149–160 (2014).
56. Corbett, L. B. et al. Elevated cosmogenic $^{26}\text{Al}/^{10}\text{Be}$ production ratio at high latitude. *Eos abstr.* C53C–0739 (2016).
57. Party, S. S. in *Proc. Ocean Drilling Program* (eds Jansen, E., Raymo, M. E. & Blum, P.) 345–387 (Ocean Drilling Program, 1996).
58. Kohl, C. P. & Nishiizumi, K. Chemical isolation of quartz for measurement of in-situ-produced cosmogenic nuclides. *Geochim. Cosmochim. Acta* **56**, 3583–3587 (1992).
59. Corbett, L. B., Bierman, P. R. & Rood, D. H. An approach for optimizing in situ cosmogenic ^{10}Be sample preparation. *Quat. Geochronol.* **33**, 24–34 (2016).
60. Rood, D. H., Brown, T. A., Finkel, R. C. & Guilderson, T. P. Poisson and non-Poisson uncertainty estimations of $^{10}\text{Be}/^9\text{Be}$ measurements at LLNL-CAMS. *Nucl. Instrum. Methods Phys. Res. B* **294**, 426–429 (2013).
61. Rood, D. H., Hall, S., Guilderson, T. P., Finkel, R. C. & Brown, T. A. Challenges and opportunities in high-precision Be-10 measurements at CAMS. *Nucl. Instrum. Methods Phys. Res. B* **268**, 730–732 (2010).
62. Xu, S., Freeman, S. P. H. T., Rood, D. H. & Shanks, R. P. Decadal ^{10}Be , ^{26}Al and ^{36}Cl QA measurements on the SUERC 5 MV accelerator mass spectrometer. *Nucl. Instrum. Methods Phys. Res. B* **361**, 39–42 (2015).
63. Cande, S. C. & Kent, D. V. Revised calibration of the geomagnetic polarity timescale for the Late Cretaceous and Cenozoic. *J. Geophys. Res.* **100**, 6093–6095 (1995).

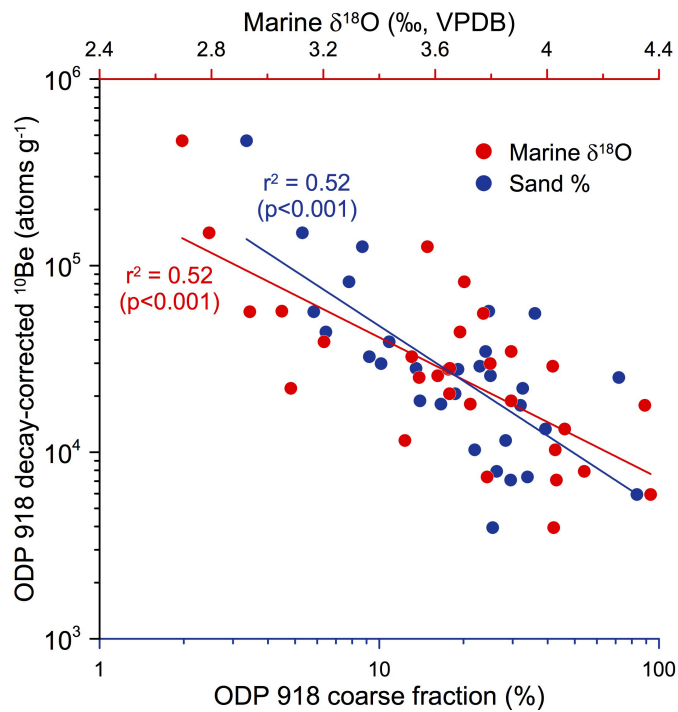
64. Fukuma, K. in *Proc. ODP Sci. Res.* Vol. 152 (eds Saunders, A. D., Larsen, H. C. & Wise, S. W. Jr) 265–269 (Ocean Drilling Program, 1998).
65. Wei, W. in *Proc. ODP Sci. Res.* Vol. 152 (eds Saunders, A. D., Larsen, H. C. & Wise, S. W. Jr) 147–160 (Ocean Drilling Program, 1998).
66. Spezzaferri, S. in *Proc. ODP Sci. Res.* Vol. 152 (eds Saunders, A. D., Larsen, H. C. & Wise, S. W. Jr) 161–189 (Ocean Drilling Program, 1998).
67. Norris, T. L., Gancarz, A. J., Rokop, D. J. & Thomas, K. W. Half-life of ^{26}Al . *J. Geophys. Res.* **88**, B331–B333 (1983).
68. Carlson, A. E. *et al.* Earliest Holocene south Greenland ice sheet retreat within its late Holocene extent. *Geophys. Res. Lett.* **41**, 5514–5521 (2014).
69. Rohling, E. J. *et al.* Sea-level and deep-sea-temperature variability over the past 5.3 million years. *Nature* **508**, 477–482 (2014).
70. de Boer, B., Lourens, L. J. & van de Wal, R. S. W. Persistent 400,000-year variability of Antarctic ice volume and the carbon cycle is revealed throughout the Plio-Pleistocene. *Nat. Commun.* **5**, 2999 (2014).
71. de Vernal, A. & Hillaire-Marcel, C. Natural variability of Greenland climate, vegetation, and ice volume during the past million years. *Science* **320**, 1622–1625 (2008).
72. Dutton, A. *et al.* Sea-level rise due to polar ice-sheet mass loss during past warm periods. *Science* **349**, 153 (2015).
73. Elderfield, H. *et al.* Evolution of ocean temperature and ice volume through the mid-Pleistocene climate transition. *Science* **337**, 704–709 (2012).
74. Sosdian, S. & Rosenthal, Y. Deep-sea temperature and ice volume changes across the Pliocene-Pleistocene climate transitions. *Science* **325**, 306–310 (2009).
75. Hansen, J., Sato, M., Russell, G. & Kharecha, P. Climate sensitivity, sea level and atmospheric carbon dioxide. *Phil. Trans. R. Soc. Lond. A* **371**, <http://dx.doi.org/10.1098/rsta.2012.0294> (2013).
76. Eldrett, J. S., Harding, I. C., Wilson, P. A., Butler, E. & Roberts, A. P. Continental ice in Greenland during the Eocene and Oligocene. *Nature* **446**, 176–179 (2007).
77. Tripathi, A. K. *et al.* Evidence for glaciation in the Northern Hemisphere back to 44 Ma from ice-rafted debris in the Greenland Sea. *Earth Planet. Sci. Lett.* **265**, 112–122 (2008).
78. Nishiizumi, K. *et al.* In situ produced cosmogenic nuclides in GISP2 rock core from Greenland summit. *Eos* **77**, abstr. OS41B–10 (1996).
79. Willerslev, E. *et al.* Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* **317**, 111–114 (2007).
80. Reyes, A. V. *et al.* South Greenland ice-sheet collapse during Marine Isotope Stage 11. *Nature* **510**, 525–528 (2014).
81. Bennike, O. *et al.* A multi-proxy study of Pliocene sediments from Île de France, North-East Greenland. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **186**, 1–23 (2002).
82. Feyling-Hanssen, R. W., Funder, S. & Petersen, K. S. The Lodin Elv Formation; a Plio-Pleistocene occurrence in Greenland. *Bull. Geol. Soc. Den.* **31**, 81–106 (1983).
83. Bennike, O. *et al.* Early Pleistocene sediments on Store Koldewey, northeast Greenland. *Boreas* **39**, 603–619 (2010).
84. Knutz, P. C., Hopper, J. R., Gregersen, U., Nielsen, T. & Japsen, P. A contourite drift system on the Baffin Bay–West Greenland margin linking Pliocene Arctic warming to poleward ocean circulation. *Geology* **43**, 907–910 (2015).
85. Koenig, S. J. *et al.* Ice sheet model dependency of the simulated Greenland Ice Sheet in the mid-Pliocene. *Clim. Past* **11**, 369–381 (2015).
86. Lunt, D. J., Foster, G. L., Haywood, A. M. & Stone, E. J. Late Pliocene Greenland glaciation controlled by a decline in atmospheric CO₂ levels. *Nature* **454**, 1102–1105 (2008).
87. Solgaard, A. M., Reeh, N., Japsen, P. & Nielsen, T. Snapshots of the Greenland ice sheet configuration in the Pliocene to early Pleistocene. *J. Glaciol.* **57**, 871–880 (2011).
88. Dolan, A. M. *et al.* Using results from the PlioMIP ensemble to investigate the Greenland Ice Sheet during the mid-Pliocene Warm Period. *Clim. Past* **11**, 403–424 (2015).



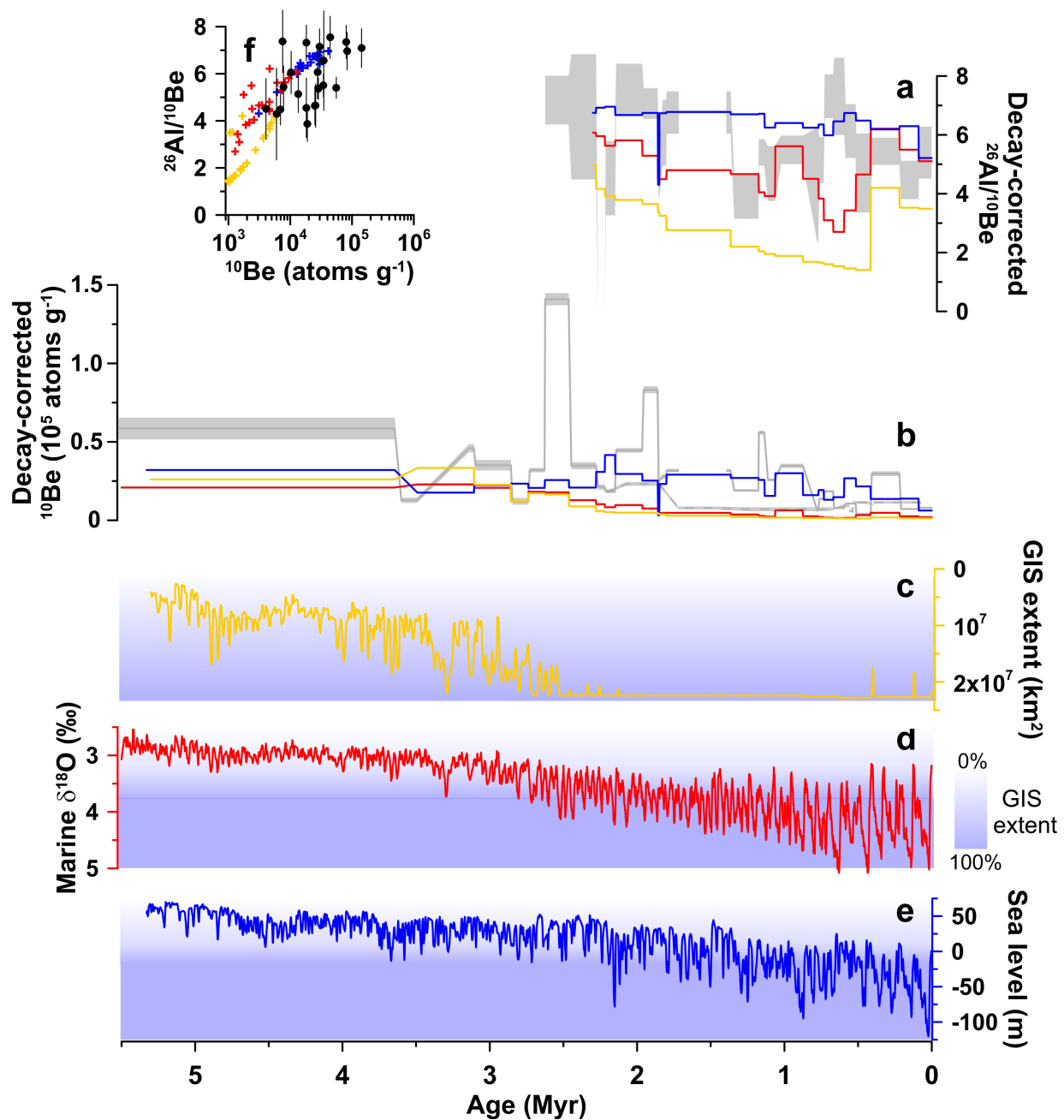
Extended Data Figure 1 | Age–depth models for Sites 918 and 987.
 Chronostratigraphic constraints^{19,40} are identified by symbols.
 (mbsf, metres below seafloor.)



Extended Data Figure 2 | Site 918 planktonic $\delta^{18}\text{O}$ stratigraphy. **a**, The global benthic LR04 $\delta^{18}\text{O}$ stack on its timescale²⁰. VPDB, Vienna Pee-Dee Belemnite standard. **b**, A planktonic (*N. pachyderma*, left-coiling) $\delta^{18}\text{O}$ record from ODP site 646 off southern Greenland, also on the global benthic $\delta^{18}\text{O}$ stack timescale⁷¹. **c**, The planktonic (*N. pachyderma*, left-coiling) $\delta^{18}\text{O}$ record from Site 918 on its depth scale. Notable interglacials in the LR04 stack and their interpreted correlatives at Site 918 are numbered, and the location of the Brunhes–Matuyama magnetic reversal in each record is denoted by the vertical dotted black line. The well resolved ODP site 646 $\delta^{18}\text{O}$ record is shown to provide a nearby planktonic record for comparison to Site 918.

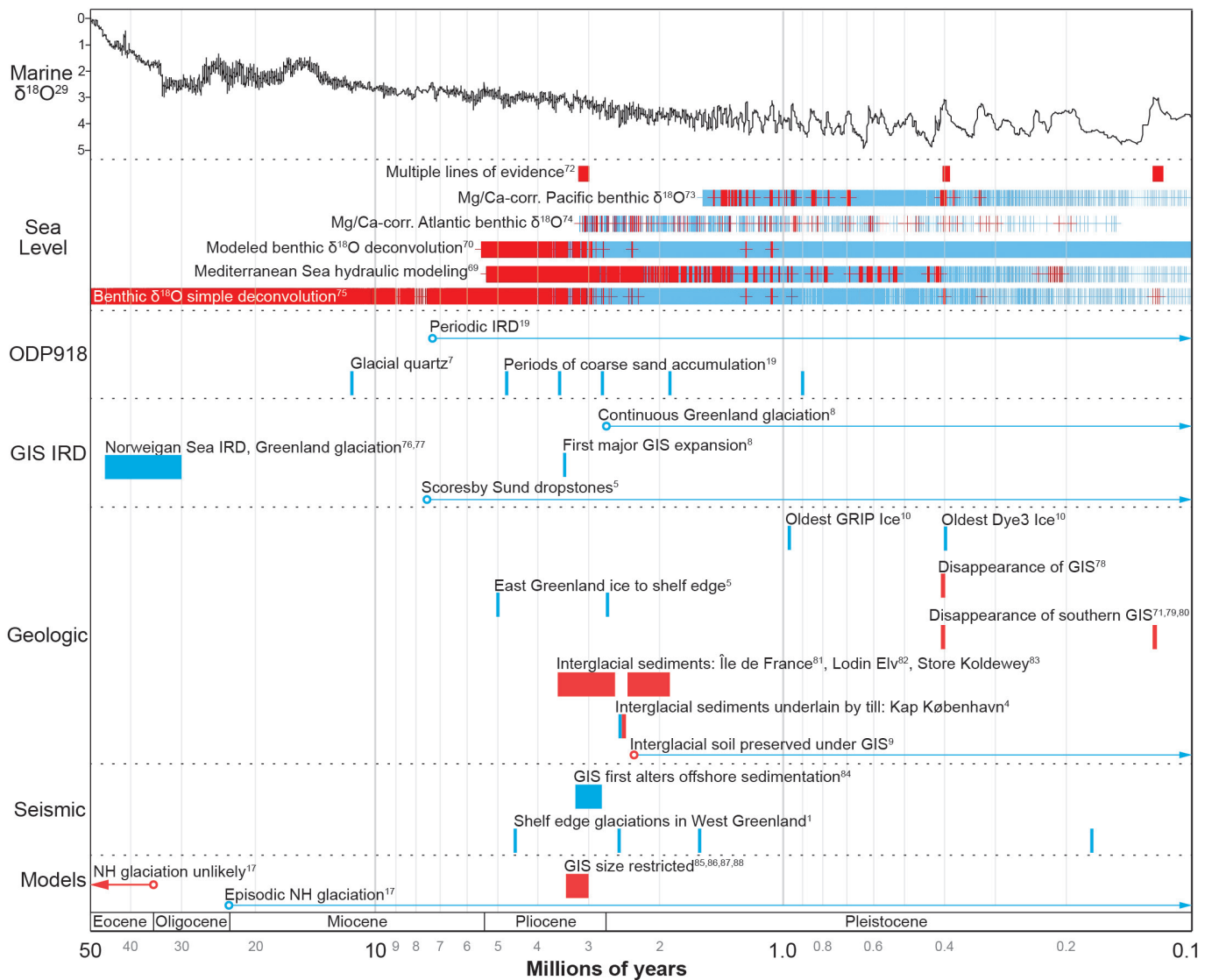


Extended Data Figure 3 | Comparing Site 918 decay-corrected ^{10}Be concentrations to Site 918 sand ($>63\ \mu\text{m}$) concentrations and marine $\delta^{18}\text{O}$ over the past 7.5 Myr. All data have been binned to the same age intervals as the ^{10}Be data. Coarse fraction indicates sand. The r^2 and P values quantify the correlations of the ^{10}Be concentrations with the sand concentrations and marine $\delta^{18}\text{O}$ values.



Extended Data Figure 4 | A simple forward model of Greenlandic cosmogenic-nuclide concentrations and ratios over the past 5 million years. a–e, Simulated (coloured lines) $^{26}\text{Al}/^{10}\text{Be}$ ratios (a) and ^{10}Be concentrations (b) of glacially eroded material from a box model with ice extent parameterized as a function of GIS extent from a full ice-sheet model⁷⁰ (c), marine $\delta^{18}\text{O}$ (ref. 20) (d), and sea level⁶⁹ (e). The colours of the simulated records in a and b correspond to the associated drivers of the

model in c, d and e. The ice extent parameterization is represented by the blue shading in c, d and e. Sites 918 and 987 cosmogenic-nuclide records are shown by 1 σ grey shading in a and b, and simulated records have been binned to the same resolution. f, $^{26}\text{Al}/^{10}\text{Be}$ – ^{10}Be relationships in the simulated (colours) and ODP Site 918 (black) records. Error bars are 1 σ . See Methods for model details and <https://github.com/shakun/Bierman-et-al-2016-Nature> for computer code.



Extended Data Figure 5 | Fully referenced version of Fig. 2. NH, Northern Hemisphere. Data are from refs 1, 4, 5, 7–10, 17, 19 and 69–88.

Dynamics of a seafloor–spreading episode at the East Pacific Rise

Yen Joe Tan¹, Maya Tolstoy¹, Felix Waldhauser¹ & William S. D. Wilcock²

Seafloor spreading is largely unobserved because 98 per cent of the global mid-ocean-ridge system is below the ocean surface. Our understanding of the dynamic processes that control seafloor spreading is thus inferred largely from geophysical observations of spreading events on land at Afar in East Africa and Iceland¹. However, these are slow-spreading centres¹ influenced by mantle plumes^{2,3}. The roles of magma pressure and tectonic stress in the development of seafloor spreading are still unclear. Here we use seismic observations to show that the most recent eruption at the fast-spreading East Pacific Rise just North of the Equator initiated at a melt-rich segment about 5 kilometres long⁴. The change in static stress then promoted almost-concurrent rupturing along at least 35 kilometres of the ridge axis, where tectonic stress had built up to a critical level, triggering magma movement. The location of impulsive seismic events indicative of lava reaching the seafloor⁵ suggests that lava subsequently erupted from multiple isolated^{6,7} magma lenses (reservoir chambers) with variable magma ascent rates, mostly within 48 hours. Therefore, even at magmatically robust fast-spreading ridges, a substantial portion of the spreading may be due to tectonic stress building up to a critical level rather than magma overpressure in the underlying magma lenses.

The global mid-ocean ridge (MOR) system is located mostly in the deep ocean. Therefore, remote detections of seismic signals at submarine ridges are interpreted by analogy to observations at subaerial ridges to infer dynamic spreading processes. Only ten seafloor spreading events have been seismically characterized to date^{5,8}, with hydroacoustic detections of migrating earthquake swarms at intermediate-spreading ridges (about 4–9 cm yr⁻¹ at the full spreading rate) interpreted as lateral dike propagations⁹, similar to observations at slow-spreading centres (<4 cm yr⁻¹) in Afar and Iceland^{10–13}. However, less is known about fast-spreading ridges (>9 cm yr⁻¹), where the relatively thin lithosphere produces earthquakes that are extremely difficult to locate with existing hydroacoustic monitoring networks¹⁴.

The most recent eruption occurring near 9° 50' N at the fast-spreading East Pacific Rise is the first observed repeat eruption at a mid-ocean ridge^{15,16}. The event was recorded by ocean-bottom seismometers (OBSs) deployed between 9° 49' N and 9° 51' N from May 2005 to April 2006, although only three seismometers survived the lava flow and unrelated instrument failure. A dike event on 22 January 2006 was inferred from an hour-long peak in seismic amplitude¹⁶ and hydroacoustically detected (T-wave) earthquakes¹⁴. However, the eruption is widely referred to as the 2005–2006 eruption and cited as having happened in multiple pulses over a period of 7–10 months^{6,7,17,18} on the basis of radiometric dating results¹⁹, similar to the previous eruption at 9° 50' N in 1991–1992 (refs 15, 20). This implies that the eruptions were modulated by pulses of magma injection into the underlying axial magma lenses (AMLs).

We analysed the OBS data collected from May 2005 to April 2006 and detected three types of seismic events—impulsive lava events, local earthquakes, and long-period events (Figs 1 and 2). The lava events have impulsive waveforms with a dominant frequency around

22 Hz (Extended Data Fig. 1), similar to those previously suggested²¹ and recently confirmed to be associated with fresh lava reaching the seafloor⁵. This association is further supported by their locations (see Methods) which coincide extremely well with the fresh lava flow boundaries (Fig. 1b) mapped using digital seafloor imagery collected in 2006–2007 (ref. 17). Their spatiotemporal evolution, where events at individual flow lobes show initial migration away from the ridge axis before concentrating at the toe of the flows (Fig. 3, Extended Data Fig. 2a), is also consistent with lava freezing over as it reaches the seafloor and forming subsurface lava channels that allow subsequent flows to propagate off-axis¹⁵. These events could be generated by explosions from magma degassing²² or the interaction of lava with seawater²³ as lava reaches the seafloor, and are consistent with evidence for explosive activity at 9° 50' N during the 1991–1992 eruption¹⁵.

These lava events were detected starting 22 January 2006 and lasted for about a week, with about 85% of the located events occurring during the first two days (Extended Data Fig. 2b). This suggests that most of the lava erupted rapidly in the first 48 h and unequivocally dates the eruption to January 2006 rather than 2005–2006. With an estimated $22 \times 10^6 \text{ m}^3$ of lava erupted¹⁷, this gives an average eruption rate of around $100 \text{ m}^3 \text{ s}^{-1}$. This is much lower than the eruption rate of $10^3 \text{ m}^3 \text{ s}^{-1}$ to $10^6 \text{ m}^3 \text{ s}^{-1}$ derived from numerical modelling²⁴ for the less voluminous 1991–1992 eruption¹⁷, despite the melt from both eruptions being derived from the same mantle source⁷. Assuming that the lower eruption rate corresponds to a lower magma driving pressure, this suggests that tectonic stress may be important in initiating the 2006 eruption.

In the two years leading up to the eruption, earthquake activity rate gradually ramped up¹⁶. The limitations of the small surviving OBS array mean that our earthquake locations are largely restricted to within a small area between 9° 49.8' N and 9° 51.1' N and from the axial summit trough (AST) to approximately 2 km west of the axis (Fig. 1c). In the hours preceding the eruption, earthquakes in this area ruptured an eastward-dipping structure that goes down to approximately the depth of the AML near 9° 50.5' N (Extended Data Fig. 3). The structure might represent a zone of developing normal faults that were activated by the build-up of magma pressure⁷ or tectonic stress²⁵ since the last eruption. The build-up of stress is also indicated by the increasing magnitude of these local earthquakes in the hour preceding the volcanic crisis (Extended Data Fig. 4), which culminated in a cluster of T-wave earthquakes of body-wave magnitudes $M_b \approx 2\text{--}3.5$ near 9° 50.5' N (Fig. 1a) that was detected by a regional hydrophone array¹⁴.

Fifteen minutes after the first T-wave earthquakes near 9° 50.5' N, a series of long-period events were detected around the same region (Fig. 2). These signals have a dominant frequency around 1.6 Hz. Long-period events are regularly observed preceding volcanic eruptions and have been suggested to originate from a resonating source in a fluid-filled conduit or crack²⁶. These events are located (see Methods) west of the ridge axis within our seismic array (Fig. 1c). We did not find any long-period events that are not detected by all stations. This suggests that similar signals are unlikely to occur just outside the array.

¹Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA. ²School of Oceanography, University of Washington, Seattle, Washington 98195, USA.

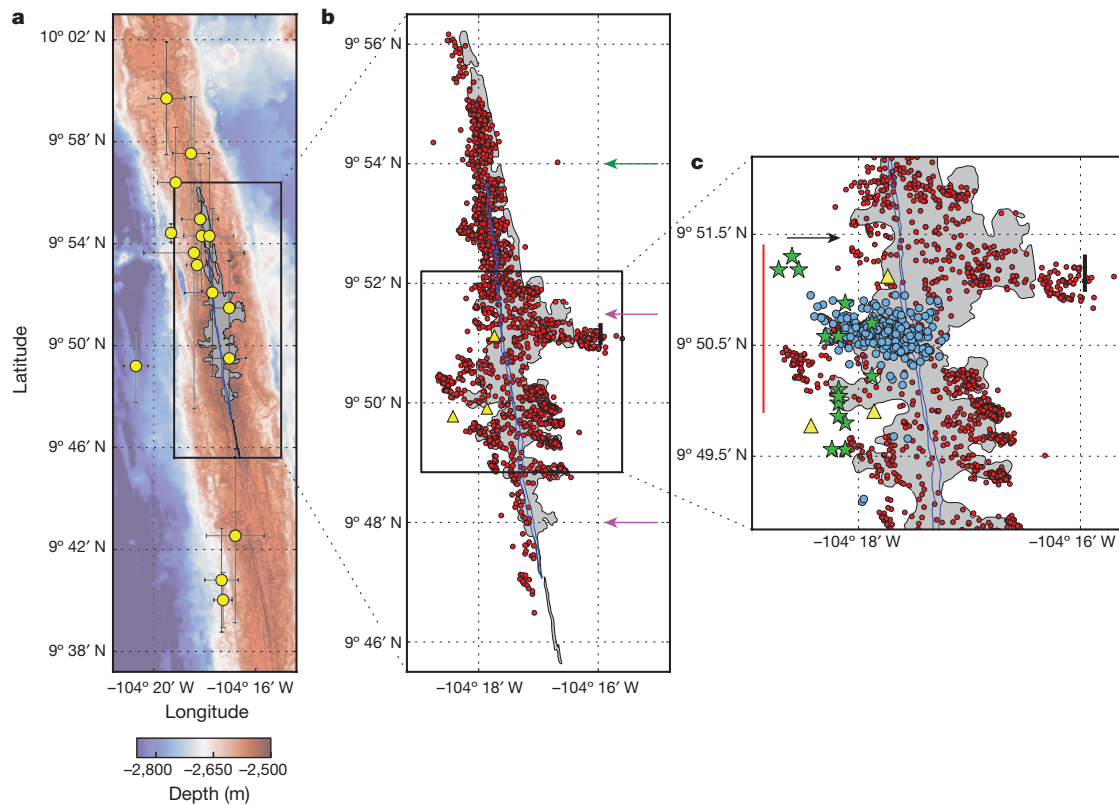


Figure 1 | Location of various seismic events. **a**, Location of T-wave earthquakes¹⁴ (yellow dots) and lava flow map¹⁷ (shaded region). Error bars represent 68% confidence interval¹⁴. **b**, Location of seismometers (yellow triangles) and impulsive lava events (red dots), showing close spatial correlation with the lava flow map¹⁷. The AST¹⁷ is shown as a thin blue line with green arrow marking its end. The bold black line marks the revised extent of the 2006 lava flow farther off-axis than originally

identified¹⁸. AML disruptions (purple arrows) were inferred from seismic reflection studies post-eruption⁶. **c**, Location of impulsive lava events (red dots), local earthquakes (blue dots), and long-period events (green stars). The black arrow marks the break in AST and the northern extent of melt-depleted sub-AML³⁰ (red line) associated with the end of the northward migration of long-period events. The bold black line marks the revised extent of the 2006 lava flow farther off-axis than originally identified¹⁸.

These long-period events are the largest-amplitude signals detected by the OBSs throughout their deployment. Seismic reflection studies after the eruption have shown melt bodies up to a kilometre off-axis in the region²⁷. Therefore, these long-period events might reflect dike initiation from the western edge of the AML, consistent with models that suggest that diking might initiate at the edge of a magma lens²⁸. The first T-wave earthquakes occurred before the long-period events, which suggests that the faults could have breached through to the AML and initiated magma movement²⁵.

The long-period events started near 9° 49.5' N and migrated about 4 km northward over the next 30 min, terminating at 9° 51.3' N near a break in the AST¹⁷ and the AML⁶ (Figs 1c and 2). These events may reflect flow instabilities as magma rises vertically from the magma lens while the AML ruptures northward (Fig. 4), or lateral dike propagation along the segment that terminated at a stress barrier²⁹. The along-axis extent of long-period events matches that of the melt-depleted section of a deeper magma lens (sub-AML) underlying the top AML (Figs 1c and 2) imaged post-eruption³⁰. This suggests that the rapid draining of this segment of the AML might have induced recharge from the underlying sub-AML³⁰ (Fig. 4). Lava first reached the seafloor in this region around 2 h after the first T-wave earthquakes (Fig. 2). This implies that magma rose from the AML starting at a depth of around 1.5 km (ref. 6) at a rate of about 0.21 m s⁻¹ (see Methods). The seismicity in this central segment peaked and stopped before the start of the long-period events swarm (Fig. 2). This suggests that the spreading episode started with faulting that triggered magma movement, and not the reverse. This is different from the 1978 Krafla event, where the main earthquake swarm started after a peak in continuous tremor¹¹, and from the 2005 Afar event, where earthquake swarms continued after a peak in sporadic tremors and ultralong-period events¹².

The T-wave earthquakes subsequently clustered to the south and north of 9° 50.5' N (Fig. 2), encompassing the entire eruption area and beyond¹⁴. This is unlikely to represent lateral dike propagation because that would require an unrealistic propagation rate of approximately 14–32 m s⁻¹ (ref. 14). Therefore, we infer that the earthquake clustering reflects almost-concurrent faulting that was probably promoted by the static stress change from the diking event near 9° 50.5' N. This suggests that the plate boundary was close to failure owing to the build-up of tectonic stress over the 14 years or so since its last spreading episode²⁵. Our observation differs from those at spreading episodes at Afar and Iceland, where dikes propagated laterally for tens of kilometres over multiple weeks^{10–13}. Although these T-wave earthquakes were the largest earthquakes in the region during this period, we were not able to identify them on the OBSs, probably owing to the high seismic noise amplitude during this period and their locations farther away and outside our OBS array (Fig. 2).

The rupturing of the ridge axis then triggered vertical magma movement from the underlying, segmented AMLs^{6,7} (Fig. 4). Immediately after lava first reached the seafloor at the central segment between 9° 49' N and 9° 51' N, impulsive lava events at the adjacent segment to the north show a northward migration of about 5 km at a rate of 1.07 m s⁻¹, terminating at the end of the AST¹⁷ (Fig. 2). This reflects either dike propagation triggered by rupturing of the ridge axis or lava flowing northward within the AST. About 45 min later, lava started erupting in the northernmost region (Fig. 2) from off-axis fissures approximately 600 m east of the AST¹⁷, fed by a different AML⁷. The northernmost region has a slower inferred magma ascent rate of about 0.14 m s⁻¹ (see Methods).

The location of the first T-wave earthquakes and long-period events suggests that the eruption initiated near 9° 50.5' N. This region has the

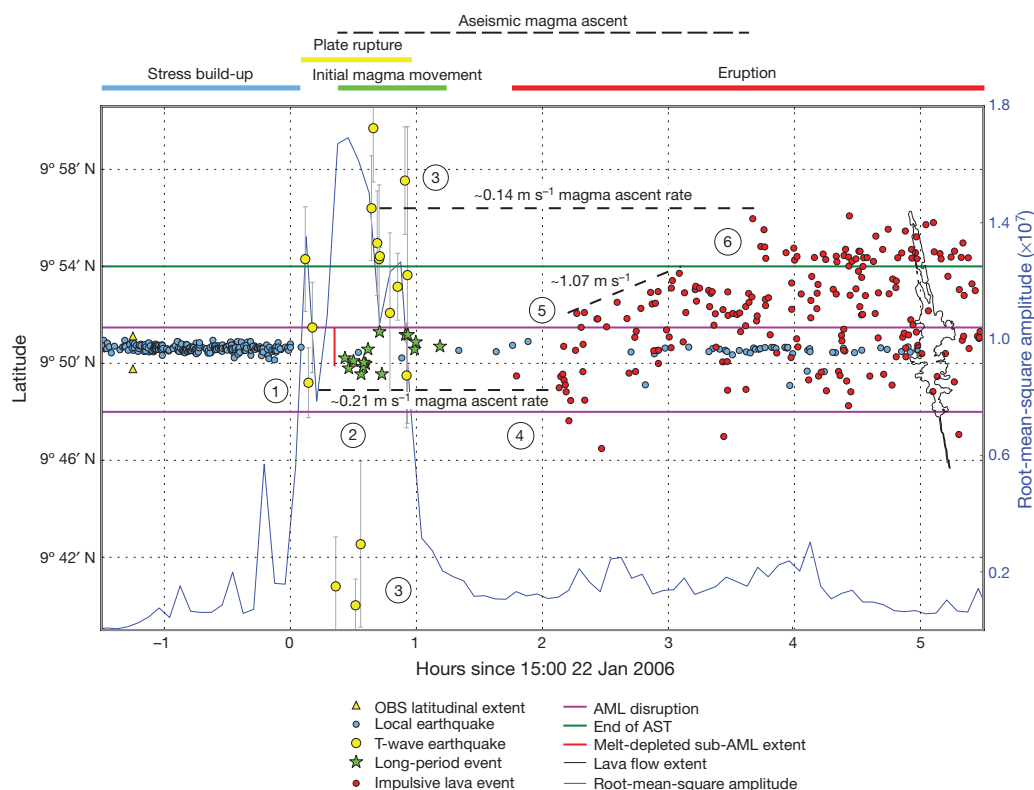


Figure 2 | Temporal and along-axis progression of various seismic events. (1) Build-up of stress culminating in plate rupture. (2) This triggered magma movement from the underlying melt-rich AML. (3) The static stress change then promoted rupturing beyond the initiation region. (4) Lava first erupted in the central segment. (5) Lava then erupted from a different AML⁶ at the adjacent segment. (6) Finally, the rupturing triggered magma from another isolated⁷ AML to rise vertically at a slower speed and erupt with a smaller volume. T-wave earthquake error bars represent the 68% confidence interval¹⁴. Magma ascent rates were calculated assuming an AML depth of 1.5 km (see Methods). AML disruptions (purple horizontal lines) and extent of melt-depleted

sub-AML (red vertical line) were inferred from seismic reflection studies post-eruption^{6,30}. New lava flow extent (solid black line), shown in map view for reference, was mapped post-eruption¹⁷. The root-mean-square amplitude of the seismic vertical channel was calculated in the 3–45-Hz band in 5-min windows (blue line). Note that flows from $\sim 9^\circ 48'$ N to $\sim 9^\circ 46'$ N did not overflow the AST and are associated with very few impulsive lava events. The southernmost T-wave earthquakes are consistent with the possible disruption of vent fluid temperatures at $9^\circ 42'$ N (ref. 17). This suggests that spreading occurred along about 35 km of the ridge axis. However, spreading could have extended beyond the distal ends of the observed T-wave earthquakes.

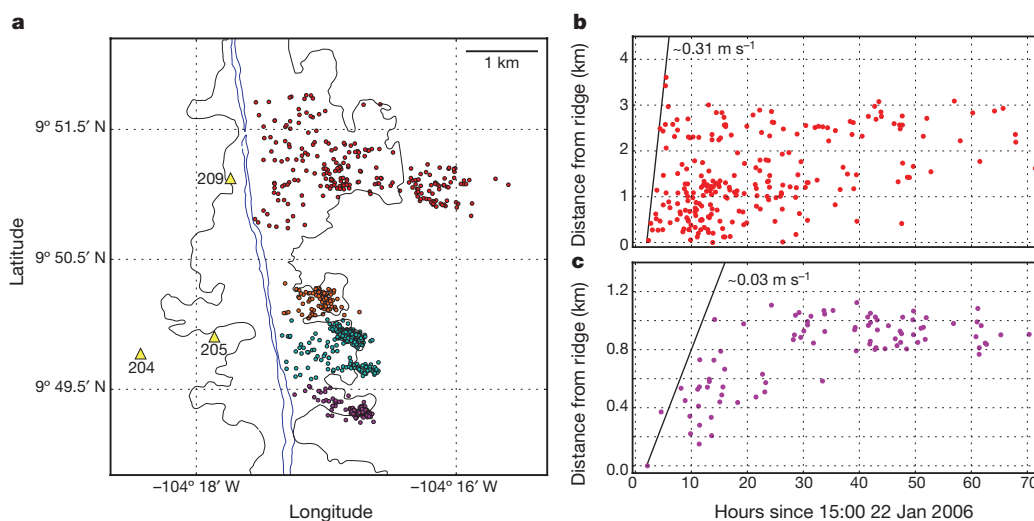


Figure 3 | Lava flow velocities. **a**, Location of seismometers (yellow triangles labelled with station numbers) and impulsive lava events (coloured dots) in the four flow lobes¹⁷ (thin black outline enclosing the lava flow). The AST¹⁷ is shown as a thin blue line. **b**, **c**, Impulsive lava events in the northernmost flow lobe propagated farther off-axis at a faster rate than events in the southernmost flow lobe. At later times, events were

concentrated at the toe (on the right) of the flows. Migration of other flow lobe events is shown in Extended Data Fig. 5. No spatiotemporal trend was observed for the flow to the west of the ridge axis, possibly because we missed many events in the first hour of the eruption (see Methods). The first events we located in the western flow lobe are already more than a kilometre off-axis (Extended Data Fig. 2a).

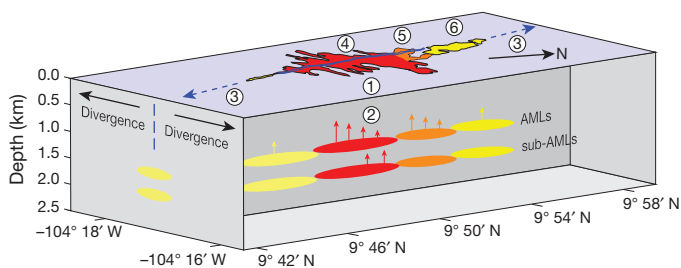


Figure 4 | Schematic representation of the dynamics of seafloor spreading at the East Pacific Rise. Numbers 1–6 mark the inferred processes represented in Fig. 2. (1) Build-up of stress culminated in plate rupture (T-wave earthquakes). (2) This triggered magma movement (long-period events) from the underlying melt-rich AML (red ellipse) which then induced recharge from a deeper sub-AML³⁰. (3) The static stress change then promoted rupturing (clustering of T-wave earthquakes) beyond the initiation region (blue dashed arrows) owing to the build-up of tectonic stress to a critical level at the plate boundary. (4) Lava first erupted (impulsive lava events) in the central segment. (5) Lava then erupted from a different AML⁶ (orange ellipse) at the adjacent segment. (6) Finally, the rupturing triggered magma from another isolated⁷ AML (yellow ellipse) to rise vertically at a slower speed and erupt with a smaller volume.

highest density of active high-temperature hydrothermal venting³¹, is near the location of the largest flow lobe from this eruption¹⁷, overlies a drained sub-AML imaged post-eruption³⁰, and has generally been considered the focus of magmatic activity and upwelling. The lavas from this region also have the highest MgO lava compositions of this eruption, suggesting relatively hot melt in the underlying AML⁷. The impulsive lava events from the flow lobe near this region propagated farthest off-axis at the fastest speed of about 0.31 m s^{-1} (Fig. 3b) compared to the events in the smaller lobes to the south that propagated at about 0.03 m s^{-1} (Fig. 3c, Extended Data Fig. 5b and c). This shows that this region had the highest lava effusion rate during the eruption, consistent with the presence of high-flow-rate morphologies¹⁸. Post-eruption seismic imaging showed that the AML underneath this 5-km-long segment is the most melt-depleted⁴, indicating that this is where most of the erupted lava was sourced from. Geochemical modelling suggests that the spreading episode was not triggered by renewed injection of magma into the underlying AML⁷, unlike at Krafla in 1978 and at Dabbahu in 2005 (ref. 1). However, the initiation at this segment may be due to long-term build-up of magma pressure⁷ or tectonic stress²⁵ since the last eruption.

Static stress change associated with the diking then promoted almost-concurrent rupturing along at least 35 km of ridge axis, triggering eruption of melt sourced from multiple AMLs^{6,7}. The less voluminous flows north of $9^{\circ} 54' \text{ N}$ and south of $9^{\circ} 48' \text{ N}$ (Fig. 1b), where lava did not even flow over the AST¹⁷, might reflect a relatively lower magma driving pressure from the underlying, isolated AMLs. Disruption in vent temperatures as far south as $9^{\circ} 42' \text{ N}$ (ref. 17) suggests that faulting may have happened even farther south without lava reaching the surface, consistent with the southernmost T-wave earthquakes on 22 January 2006.

The variable along-axis magma ascent rates and erupted lava volumes support that the multiple underlying AMLs are segmented⁶ and thus unlikely to be concurrently critically stressed from build-up of magma pressure. Instead, the build-up of tectonic stress to a critical level probably allowed almost-concurrent faulting along at least 35 km of the ridge segment, which then triggered magma movement. This is consistent with a model in which the AMLs are surrounded by hot asthenosphere and hence too weak to maintain large magma overpressure or support large elastic stress differences from the low strain rate of plate pull. The high strain rate of faulting is thus needed to breach through to the AMLs to trigger magma movement²⁵. Therefore, we can conclude that spreading along at least 30 km of the 35-km-long ridge segment (>85%) happened owing to build-up of tectonic stress.

If plate pull dominates at this magmatically robust fast-spreading ridge, it could also dominate at intermediate- and slow-spreading ridges, which generally have relatively lower magma supplies. Therefore, while MORs are often described as volcanoes that erupt owing to build-up of magma pressure^{7,19}, our study demonstrates that eruptions at MORs are a fundamentally different process largely controlled by tectonic stress due to plate pull. MORs could be viewed less as volcanoes and more as tears in the crust where magma reaches the surface when plates are pulled apart.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 March; accepted 26 September 2016.

Published online 14 November 2016.

- Wright, T. J. *et al.* Geophysical constraints on the dynamics of spreading centres from rifting episodes on land. *Nat. Geosci.* **5**, 242–250 (2012).
- Ebinger, C. J. & Sleep, N. H. Cenozoic magmatism throughout east Africa resulting from impact of a single plume. *Nature* **395**, 788–791 (1998).
- Wolfe, C. J., Bjarnason, I. T. H., Vandecar, J. C. & Solomon, S. C. Seismic structure of the Iceland mantle plume. *Nature* **385**, 245–247 (1997).
- Xu, M. *et al.* Variations in axial magma lens properties along the East Pacific Rise ($9^{\circ} 30' \text{ N}$ – $10^{\circ} 00' \text{ N}$) from swath 3-D seismic imaging and 1-D waveform inversion. *J. Geophys. Res.* **119**, 2721–2744 (2014).
- Wilcock, W. S. D., Tolstoy, M., Garcia, C., Tan, Y. J. & Waldhauser, F. Live from the seafloor: seismic signals associated with the 2015 eruption of Axial Seamount. *Eos* **OS41B-07** (2015).
- Carbotte, S. M. *et al.* Fine-scale segmentation of the crustal magma reservoir beneath the East Pacific Rise. *Nat. Geosci.* **6**, 866–870 (2013).
- Goss, A. R. *et al.* Geochemistry of lavas from the 2005–2006 eruption at the East Pacific Rise, $9^{\circ} 46' \text{ N}$ – $9^{\circ} 56' \text{ N}$: implications for ridge crest plumbing and decadal changes in magma chamber compositions. *Geochem. Geophys. Geosyst.* **11**, Q05T09 (2010).
- Tolstoy, M. Mid-ocean ridge eruptions as a climate valve. *Geophys. Res. Lett.* **42**, 1346–1351 (2015).
- Dziak, R. P. *et al.* Rapid dike emplacement leads to eruptions and hydrothermal plume release during seafloor spreading events. *Geology* **35**, 579–582 (2007).
- Abdallah, A. *et al.* Relevance of Afar seismicity and volcanism to the mechanics of accreting plate boundaries. *Nature* **282**, 17–23 (1979).
- Einarsson, P. & Brandsdóttir, B. Seismological evidence for lateral magma intrusion during the July 1978 deflation of the Krafla volcano in NE Iceland. *J. Geophys.* **47**, 160–165 (1980).
- Ayele, A. *et al.* September 2005 mega-dike emplacement in the Manda-Harrano nascent oceanic rift (Afar depression). *Geophys. Res. Lett.* **36**, L20306 (2009).
- Sigmundsson, F. *et al.* Segmented lateral dyke growth in a rifting event at Bárðarbunga volcanic system, Iceland. *Nature* **517**, 191–195 (2014).
- Dziak, R. P. *et al.* January 2006 seafloor-spreading event at $9^{\circ} 50' \text{ N}$, East Pacific Rise: ridge dike intrusion and transform fault interactions from regional hydroacoustic data. *Geochem. Geophys. Geosyst.* **10**, Q06T06 (2009).
- Haymon, R. M. *et al.* Volcanic eruption of the mid-ocean ridge along the East Pacific Rise crest at $9^{\circ} 45' \text{ N}$ – $52' \text{ N}$: direct submersible observations of seafloor phenomena associated with an eruption event in April, 1991. *Earth Planet. Sci. Lett.* **119**, 85–101 (1993).
- Tolstoy, M. *et al.* A sea-floor spreading event captured by seismometers. *Science* **314**, 1920–1922 (2006).
- Soule, S. A., Fornari, D. J., Perfit, M. R. & Rubin, K. H. New insights into mid-ocean ridge volcanic processes from the 2005–2006 eruption of the East Pacific Rise, $9^{\circ} 46' \text{ N}$ – $9^{\circ} 56' \text{ N}$. *Geology* **35**, 1079–1082 (2007).
- Fundis, A. T., Soule, S. A., Fornari, D. J. & Perfit, M. R. Paving the seafloor: Volcanic emplacement processes during the 2005–2006 eruptions at the fast spreading East Pacific Rise, $9^{\circ} 50' \text{ N}$. *Geochem. Geophys. Geosyst.* **11**, Q08O24 (2010).
- Rubin, K. H. *et al.* Volcanic eruptions in the deep sea. *Oceanography (Wash. D.C.)* **25**, 142–157 (2012).
- Rubin, K. H., Macdougall, J. D. & Perfit, M. R. ^{210}Po – ^{210}Pb dating of recent volcanic eruptions on the sea floor. *Nature* **368**, 841–844 (1994).
- Schlindwein, V., Muller, C. & Jokat, W. Seismoacoustic evidence for volcanic activity on the ultraslow spreading Gakkel Ridge, Arctic Ocean. *Geophys. Res. Lett.* **32**, L18306 (2005).
- Vergnolle, S., Brandeis, G. & Mareschal, J. C. Strombolian explosions. 2. Eruption dynamics determined from acoustic measurements. *J. Geophys. Res.* **101**, 20449–20466 (1996).
- Perfit, M. R. *et al.* Interaction of sea water and lava during submarine eruptions at mid-ocean ridges. *Nature* **426**, 62–65 (2003).
- Gregg, T. K. P., Fornari, D. J., Perfit, M. R., Haymon, R. M. & Fink, J. H. Rapid emplacement of a mid-ocean ridge lava flow on the East Pacific Rise at $9^{\circ} 46' \text{ N}$ – $51' \text{ N}$. *Earth Planet. Sci. Lett.* **144**, E1–E7 (1996).
- Buck, W. R. On the frequency of dike intrusion episodes at spreading centers. *Eos* **T41B-1566** (2006).
- Chouet, B. A. Long-period volcano seismicity: its source and use in eruption forecasting. *Nature* **380**, 309–316 (1996).

27. Canales, J. P. *et al.* Network of off-axis melt bodies at the East Pacific Rise. *Nat. Geosci.* **5**, 279–283 (2012).
28. Bohnenstiehl, D. R. & Carbotte, S. M. Faulting patterns near 19°30'S on the East Pacific Rise: fault formation and growth at a superfast spreading center. *Geochem. Geophys. Geosyst.* **2**, 1056 (2001).
29. Buck, W. R., Einarsson, P. & Brandsdottir, B. Tectonic stress and magma chamber size as controls on dike propagation: constraints from the 1975–1984 Krafla rifting episode. *J. Geophys. Res.* **111**, B12404 (2006).
30. Marjanović, M. *et al.* A multi-sill magma plumbing system beneath the axis of the East Pacific Rise. *Nat. Geosci.* **7**, 825–829 (2014).
31. Fornari, D. *et al.* in *The Thermal Structure of the Ocean Crust and the Dynamics of Hydrothermal Circulation* Vol. 148, 187–218 (AGU Geophys. Monogr., American Geophysical Union, 2004).

Acknowledgements Y.J.T. thanks W. R. Buck for discussions and reading of the manuscript. Y.J.T. thanks A. H. Barclay and R. S. Matoza for discussions.

We thank R. P. Dziak and D. R. Bohnenstiehl for providing the T-wave earthquake catalogue. We thank the captains, crews and science parties of the RV *Knorr* and RV *Atlantis*. This work was supported by NSF grant OCE-0961594.

Author Contributions Y.J.T., advised by M.T. and F.W., analysed the seismic data and wrote the manuscript. M.T. and F.W. conceived and ran the experiment. W.S.D.W.'s identification of impulsive lava signals at Axial inspired this study. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.J.T. (yjt@ldeo.columbia.edu).

Reviewer Information *Nature* thanks J. Karson and V. Schlindwein for their contribution to the peer review of this work.

METHODS

Impulsive lava event detection and location. We used a standard detection algorithm based on short-term average to long-term average of the seismic energy recorded by the seismometer, covering May 2005 to April 2006. We then manually went through the entire period to identify missed events and to pick the arrivals. No events were identified before 22 January 2006. During the first hour of the eruption, we could not locate all the events because many of them were occurring simultaneously at short distances from the stations, thus observed phases could not be uniquely assigned to individual events.

Events were located by assuming the signals were generated on the seafloor and had bounced off the sea surface once before reaching the seismometers (Extended Data Fig. 2b inset). We used an iterative least-squares method assuming a constant sound velocity of 1.5 km s^{-1} with water depth from a high-resolution bathymetry map acquired in November 2005 (ref. 32). The sound velocity profile with depth in this region varies from 1.48 km s^{-1} to 1.52 km s^{-1} (ref. 33). We quantify our location uncertainties due to a simplified sound velocity model by locating the events with a sound velocity of 1.48 km s^{-1} and 1.52 km s^{-1} . This shifts our original catalogue locations by an average of 107 m. We quantify our location uncertainties due to pick error by adding normally distributed noise of standard deviation 0.1 s to the arrival times before locating the events. This is performed on the whole catalogue 100 times. On average, the event locations are shifted by $<1 \text{ m}$. This small uncertainty may be explained by the slow sound velocity, as the locations are well constrained by the station relative arrival times (on the order of seconds, which is much larger than the estimated pick error of 0.1 s). Event magnitudes were calculated as the logarithm of the displacement (in micrometres) multiplied by the signal travel distance (in kilometres). This assumes no attenuation other than from geometric spreading.

Earthquake detection and location. We manually went through a 6.5-h period starting from approximately 3 h before the first impulsive lava event was identified and handpicked the P- and S-wave arrivals. This was the period of highest seismic root-mean-square amplitude during the deployment¹⁶. These events were then used as templates for an array-based waveform correlation detection algorithm³⁴. We then manually reviewed the detected events. Events with both P- and S-wave arrivals at all three stations were located using the NonLinLoc grid-search program³⁵. A window of 0.5 s around each arrival was extracted for cross-correlation to get differential arrival times. Only differential arrival times with cross-correlation coefficient above 0.7 were used in the relocation using the hypoDD program³⁶. Depth-dependent P- and S-wave one-dimensional velocity models were used³⁷. Local magnitudes (M_L) were derived using the formulation described in ref. 38. Synthetic test suggests that the eastward dipping structure we observed is resolvable (Extended Data Fig. 6). Least-squares errors for the relative locations were computed³⁶ using a subset of better-constrained events (template events and events located within the OBS array). The mean relative location errors are 120 m for longitude, 40 m for latitude, and 220 m for depth (Extended Data Fig. 7).

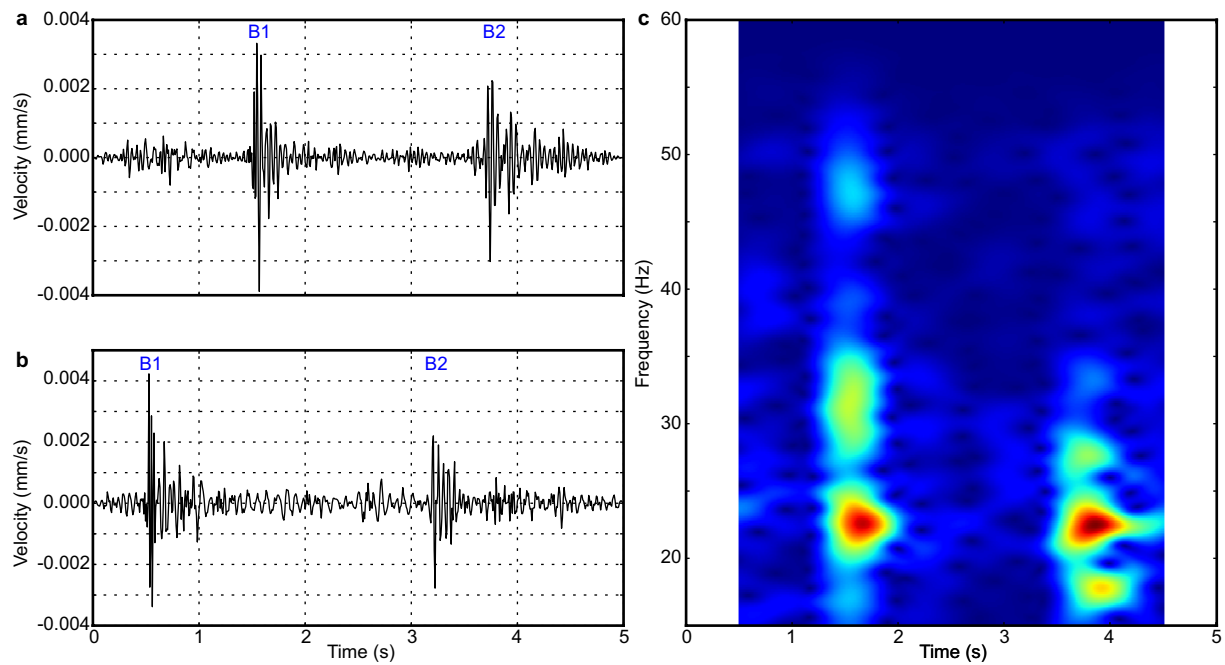
Long-period event detection and location. We manually picked the first arrivals and located the events using a grid search method that minimizes the standard deviation of predicted source origin times based on arrival times at each station³⁹, assuming a fixed source depth and constant velocity of 2.5 km s^{-1} . A 100-m grid spacing was used. The epicentres of the events do not change very much between the assumed depths of 0 km to 1.5 km (Extended Data Fig. 8a), which is the approximate depth of the AML⁶. Velocities between 0.5 km s^{-1} and 6.5 km s^{-1} were tested, with the velocity of 2.5 km s^{-1} giving the overall minimum standard deviation

(Extended Data Fig. 9). This velocity is consistent with a mean crustal S-wave velocity above 1.5 km depth³⁷. However, the epicentres of the events do not change very much with change in velocity assumption (Extended Data Fig. 8b), because the northward migration of the event locations is constrained by the station relative arrival time (Extended Data Fig. 10).

Magma ascent rates calculation. For the northernmost segment, the AML depth varies from 1.5 km to 1.65 km (ref. 6). The time gap between the first T-wave earthquake and the first lava event in this region is about 181 min. Therefore, the calculated magma ascent rate is between 0.14 m s^{-1} and 0.15 m s^{-1} . For the central segment, the AML depth varies from 1.45 km to 1.6 km (ref. 6). The time gap between the first T-wave earthquake and the first cluster of lava events in this region is about 117 min (Fig. 2). Therefore, the calculated magma ascent rate is between 0.21 m s^{-1} and 0.23 m s^{-1} . However, there is a single lava event that happened about 97 min after the first T-wave earthquake in this region (Fig. 2). If this event time is used instead, the calculated magma ascent rate for the central segment is between 0.25 m s^{-1} and 0.28 m s^{-1} . This event time represents a conservative estimate for the earliest lava event in the central segment because, at the start of the eruption, many events were occurring simultaneously at short distances from the stations, so that observed phases could not be uniquely assigned to individual events. The time gap between the last long-period event and the first lava event is around 36 min. If this aseismic period is the time of magma ascent instead, it would give the fastest rate of between 0.67 m s^{-1} and 0.74 m s^{-1} .

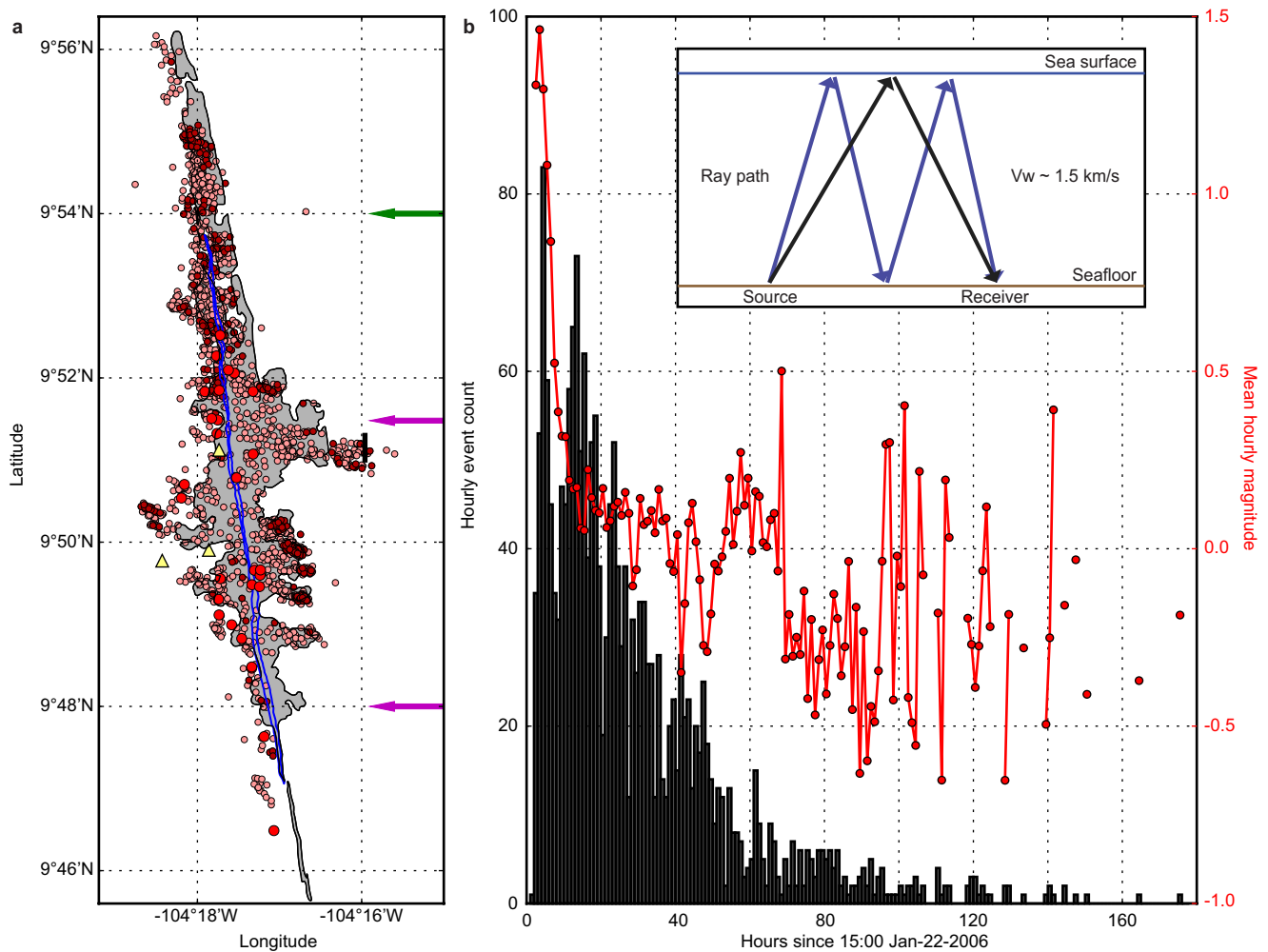
Data availability. Source data for Figs 1b and c, 2 and 3 and Extended Data Figs 2–5 are provided with the online version of the paper. All other data supporting the findings of this study are available from the corresponding author upon request.

32. White, S. M., Haymon, R. M. & Carbotte, S. A new view of ridge segmentation and near-axis volcanism at the East Pacific Rise, 8° – 12° N, from EM300 multibeam bathymetry. *Geochem. Geophys. Geosyst.* **7**, Q12005 (2006).
33. Munk, W. H., Worcester, P. & Wunsch, C. The forward problem: range-independent. *Ocean Acoust. Tomogr.* 30–114 (1995).
34. Gibbons, S. J. & Ringdal, F. The detection of low magnitude seismic events using array-based waveform correlation. *Geophys. J. Int.* **165**, 149–166 (2006).
35. Lomax, A., Michelini, A. & Curtis, A. Earthquake location, direct, global-search methods in *Encyclopedia of Complexity and System Science* Vol. 1, 2449–2473 (Springer, 2009).
36. Waldhauser, F. & Ellsworth, W. L. A double-difference earthquake location algorithm: Method and application to the Northern Hayward Fault, California. *Bull. Seismol. Soc. Am.* **90**, 1353–1368 (2000).
37. Waldhauser, F. & Tolstoy, M. Seismogenic structure and processes associated with magma inflation and hydrothermal circulation beneath the East Pacific Rise at $9^{\circ}50'$ N. *Geochem. Geophys. Geosyst.* **12**, Q08T10 (2011).
38. Bohnenstiehl, D. R., Waldhauser, F. & Tolstoy, M. Frequency-magnitude distribution of microearthquakes beneath the $9^{\circ}50'$ N region of the East Pacific Rise, October 2003 through April 2004. *Geochem. Geophys. Geosyst.* **9**, Q10T03 (2008).
39. Gottschämer, E. & Surono, I. Locating tremor and shock sources recorded at Bromo Volcano. *J. Volcanol. Geotherm. Res.* **101**, 199–209 (2000).
40. Saal, A. E., Hauri, E. H., Langmuir, C. H. & Perfit, M. R. Vapour undersaturation in primitive mid-ocean-ridge basalt and the volatile content of Earth's upper mantle. *Nature* **419**, 451–455 (2002).
41. Shinohara, H. Excess degassing from volcanoes and its role on eruptive and intrusive activity. *Rev. Geophys.* **46**, RG4005 (2008).
42. Soule, S. A. *et al.* CO_2 variability in mid-ocean ridge basalts from syn-emplacement degassing: constraints on eruption dynamics. *Earth Planet. Sci. Lett.* **327–328**, 39–49 (2012).



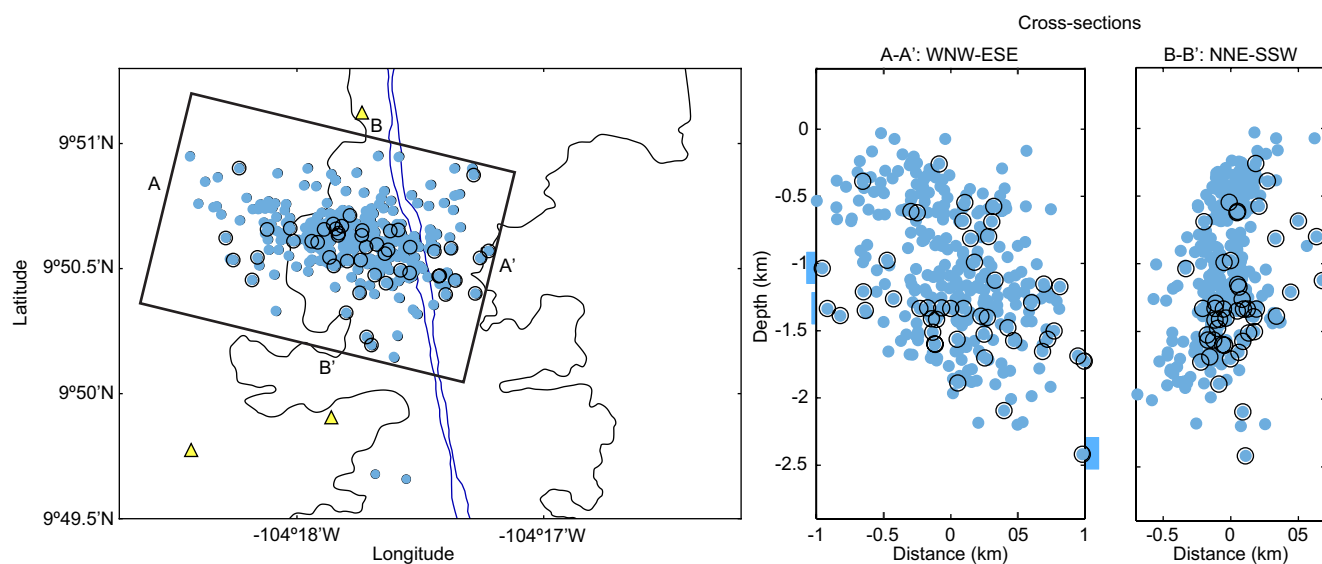
Extended Data Figure 1 | Impulsive lava event signals. **a, b,** Velocity seismograms of the same event recorded at two different stations. The difference in station relative arrival time of >1 s for the first arrival suggests a waterborne arrival (that is, the seismic wave travelled through the water column rather than through the crust), since these stations are around 2 km apart, with water velocity being 1.5 km s^{-1} and P-wave

crustal velocity being around $4\text{--}5 \text{ km s}^{-1}$ (ref. 37). The event also has multiple arrivals: the first arrival represents a signal that has bounced off the sea surface once (B1) and the second arrival represents a signal that has bounced off the sea surface twice (B2) (Extended Data Fig. 2b). **c,** The signal shows a dominant frequency of about 22 Hz.



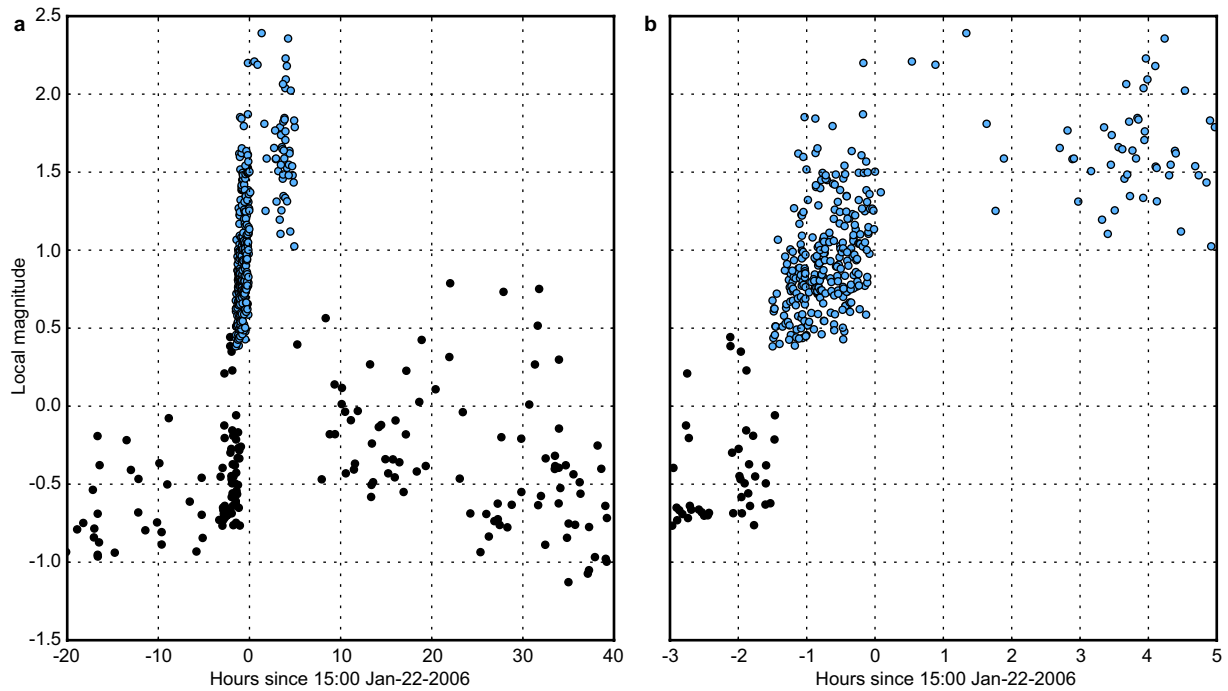
Extended Data Figure 2 | Impulsive lava events. **a**, Location of seismometers (yellow triangles) and impulsive lava events in the first hour (red), first two days (pale red), and last five days (dark red). The AST¹⁷ is shown as a thin blue line with green arrow marking its end. The bold black line marks revised 2006 lava flow farther off-axis than originally identified¹⁸. AML disruptions (purple arrows) were inferred from seismic reflection studies post-eruption⁶. **b**, Histogram of hourly impulsive lava event rate. The red line shows hourly average impulsive lava event magnitude. Mid-ocean-ridge basalt on the East Pacific Rise has been found to have low volatile content⁴⁰. Therefore, the larger-magnitude impulsive lava events during the first 5 h might be driven by excess degassing of volatiles from a larger reservoir of unerupted magma⁴¹, consistent with the estimate that <15% of the available magma in the AML was erupted¹⁷.

The lava has also been suggested to have degassed from supersaturated conditions owing to a rapid magma ascent rate from depth⁴². The breaks in the red line are due to hours when there are no events located, and hence no average magnitude data point. Event magnitudes were calculated as the logarithm of the displacement (in micrometres) multiplied by the signal travel distance (in kilometres), which is different from earthquake magnitude. The inset shows the model of the signal travel path used in locating the impulsive lava events (V_w is the velocity of the seismic wave travelling through the water column). A direct waterborne arrival is not recorded, probably because of the upward refraction of energy due to increasing velocity with water depth. The lack of crustal arrival indicates a small absolute magnitude.



Extended Data Figure 3 | Earthquake locations. The A–A' depth slice shows a 1-km-thick east-southeast (ESE)-dipping structure towards the ridge axis. The B–B' depth slice shows that the structure is subvertical in this direction (that is, north-northeast to south-southwest, NNE–SSW)

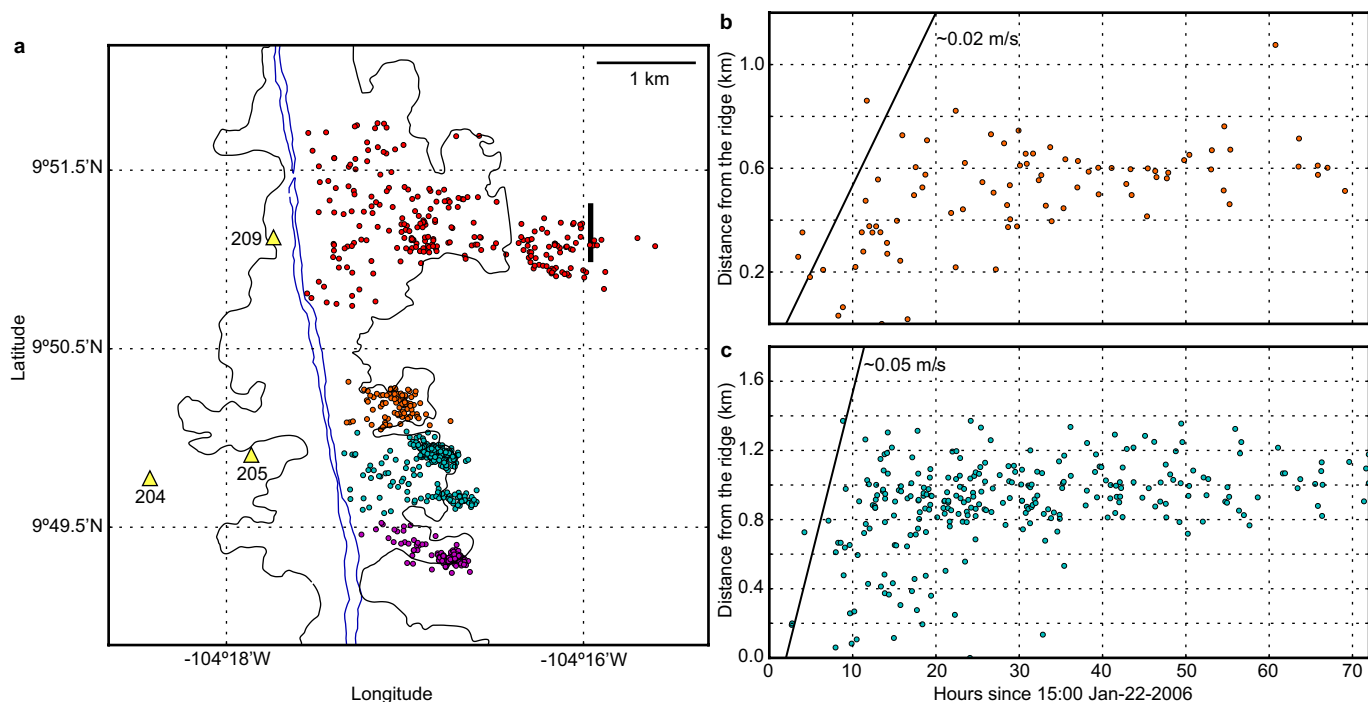
with a width of around 500 m. Earthquakes occurred between 13:30 and 20:00 on 22 January 2006, mostly in the first two hours. The 50 largest-magnitude earthquakes are highlighted with black circles.



Extended Data Figure 4 | Local magnitude of earthquakes.

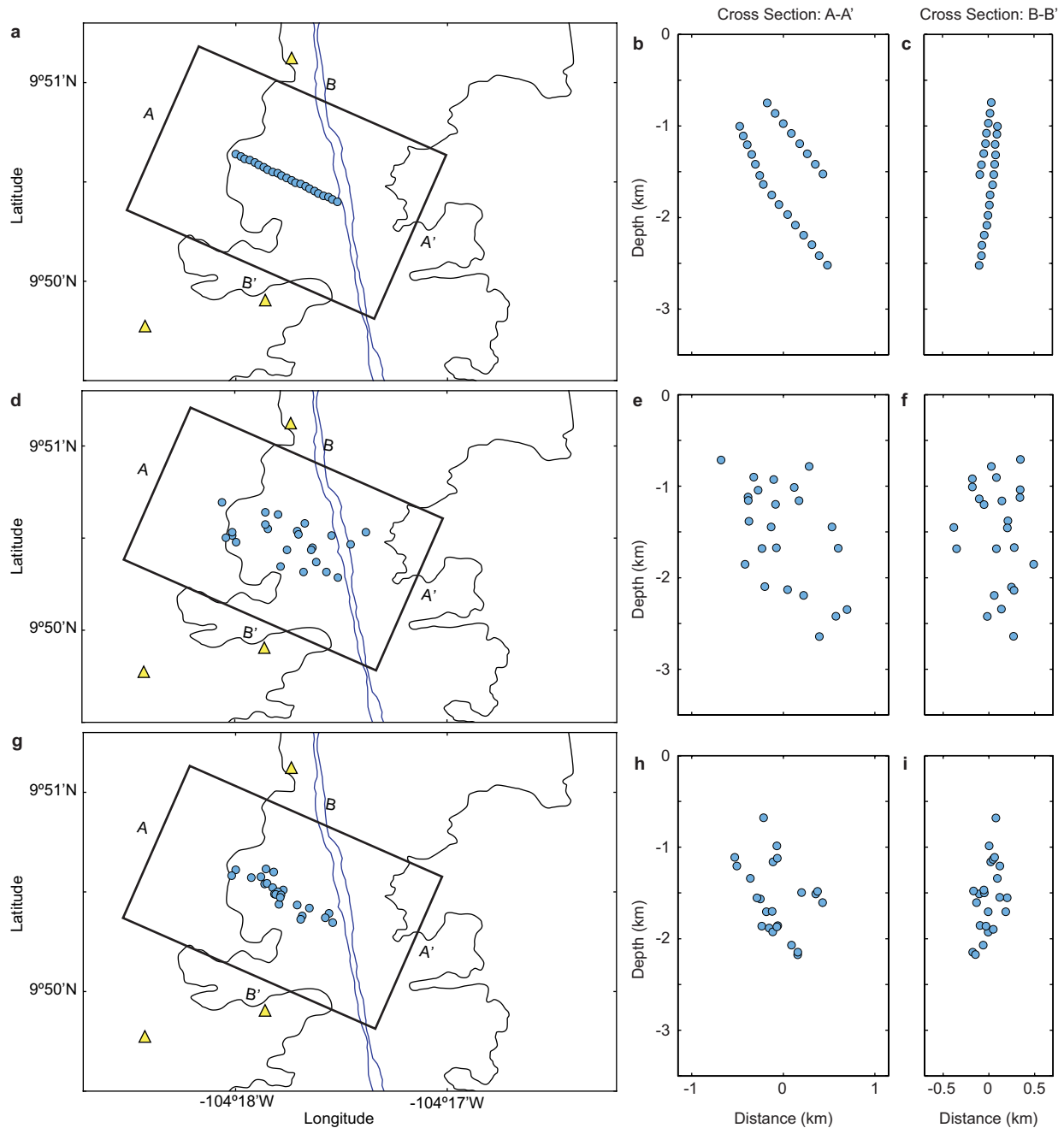
a, Earthquakes show increasing local magnitude M_L with time in a two-hour period leading up to the eruption. Blue dots represent earthquakes during the eruption period analysed in this study in detail while black dots

represent earthquakes over a longer period. **b**, An 8-h zoom-in on panel **a**, demonstrating a clear trend of increasing magnitude immediately before plate rupture.



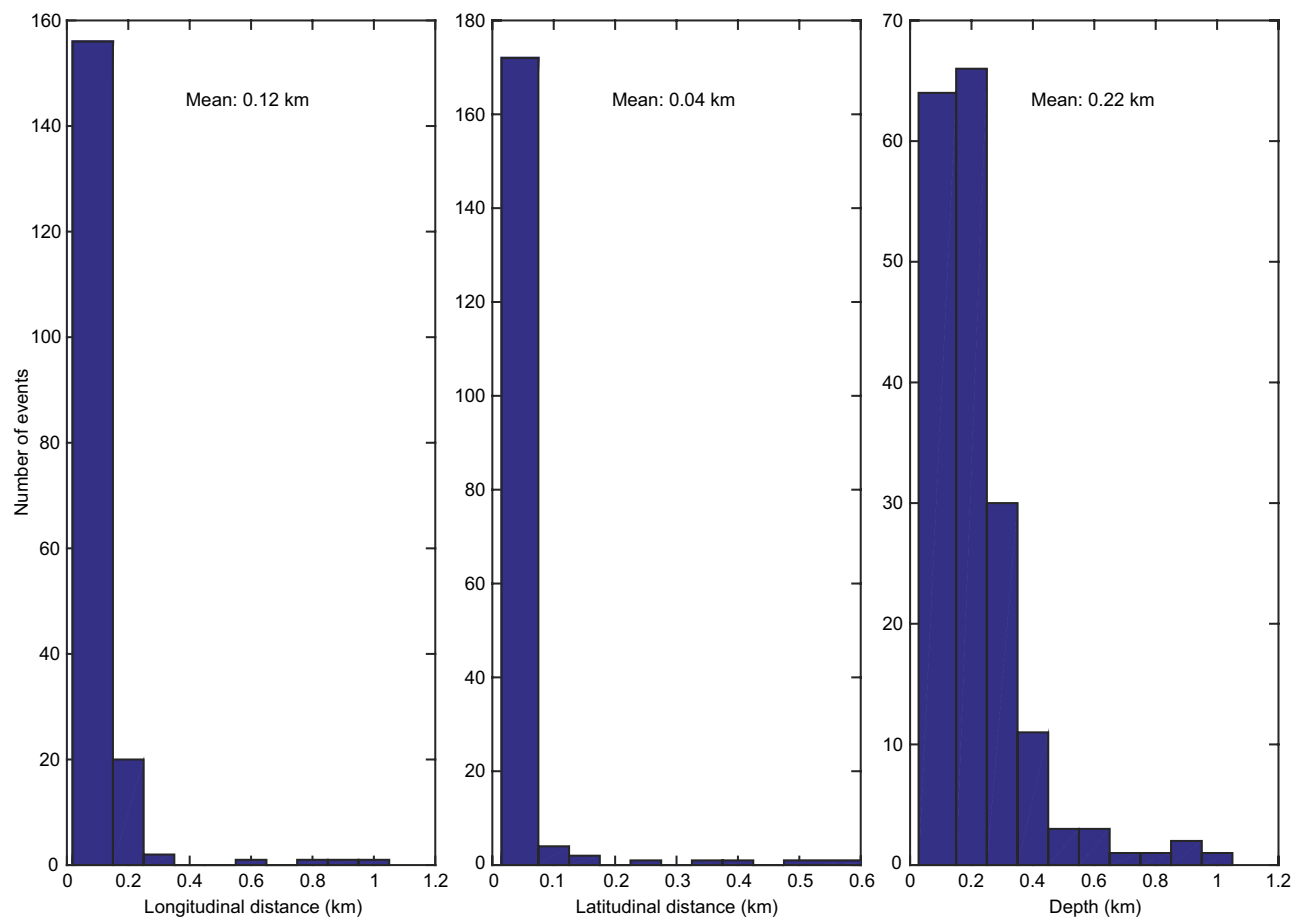
Extended Data Figure 5 | Lava flow velocities. a, Location of seismometers (yellow triangles labelled with station numbers) and impulsive lava events (coloured dots) in the four flow lobes¹⁷ (thin black outline enclosing the lava flow). The bold black line marks the revised

extent of the 2006 lava flow farther off-axis than originally identified¹⁸ (see legend to Fig. 1). b, c, Propagation of impulsive lava events off-axis in two flow lobes.

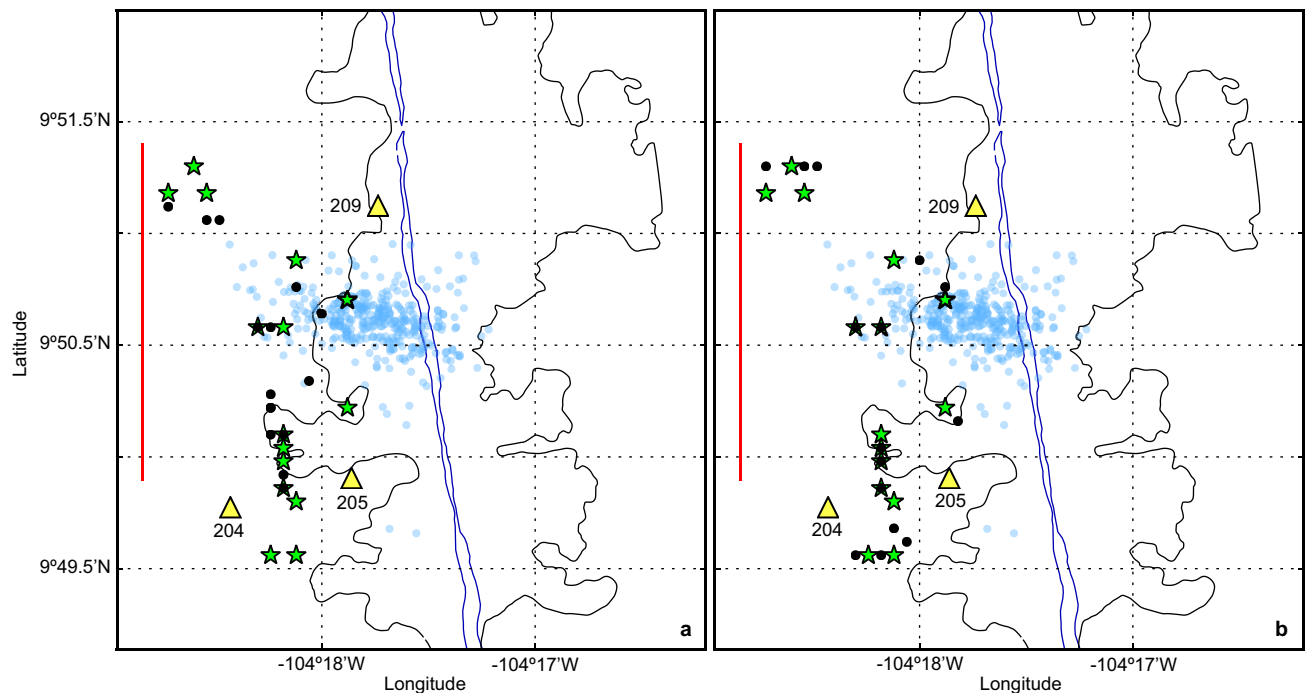


Extended Data Figure 6 | Synthetic test of earthquake locations. **a–c**, Actual locations of synthetic sources. **d–f**, Locations of synthetic sources with normally distributed noise of standard deviation 200 m added. These are used as the initial locations for double difference

relocation. **g–i**, Relocated locations of synthetic sources. Events were relocated using synthetic arrival time with normally distributed noise of standard deviation 0.01 s added.

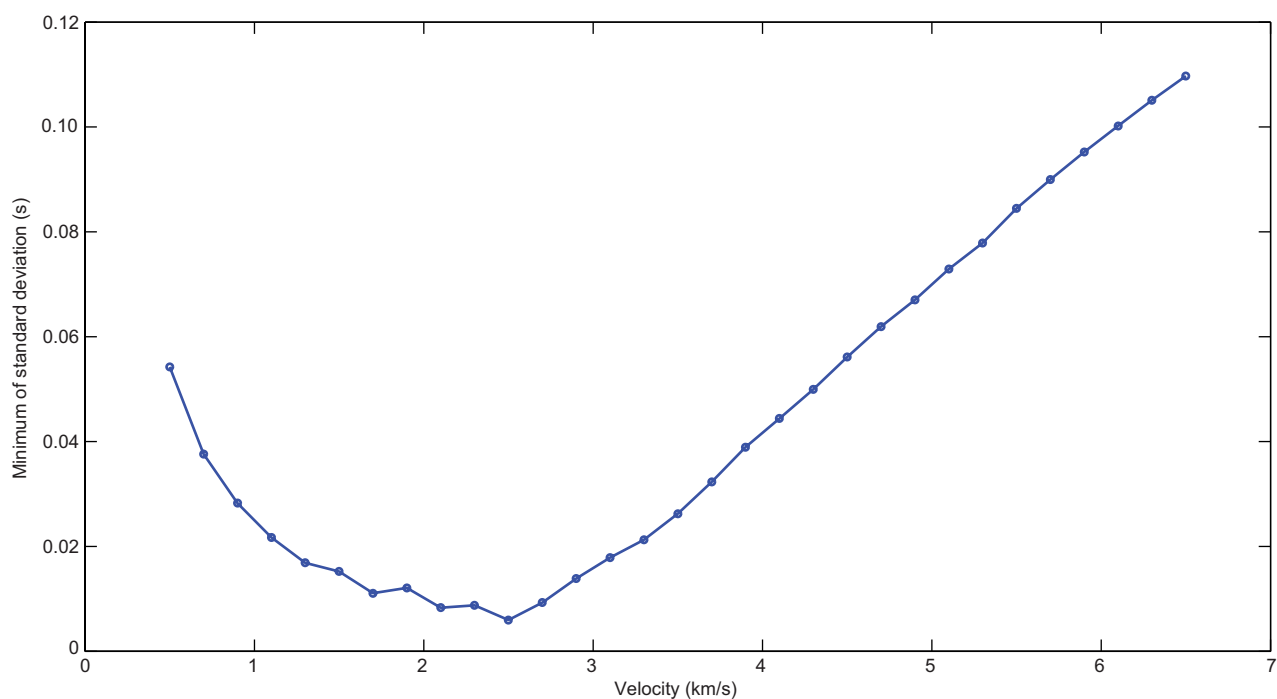


Extended Data Figure 7 | Earthquake relative location errors. Distribution of least-squares errors for the relative locations in longitudinal distance, latitudinal distance and depth.

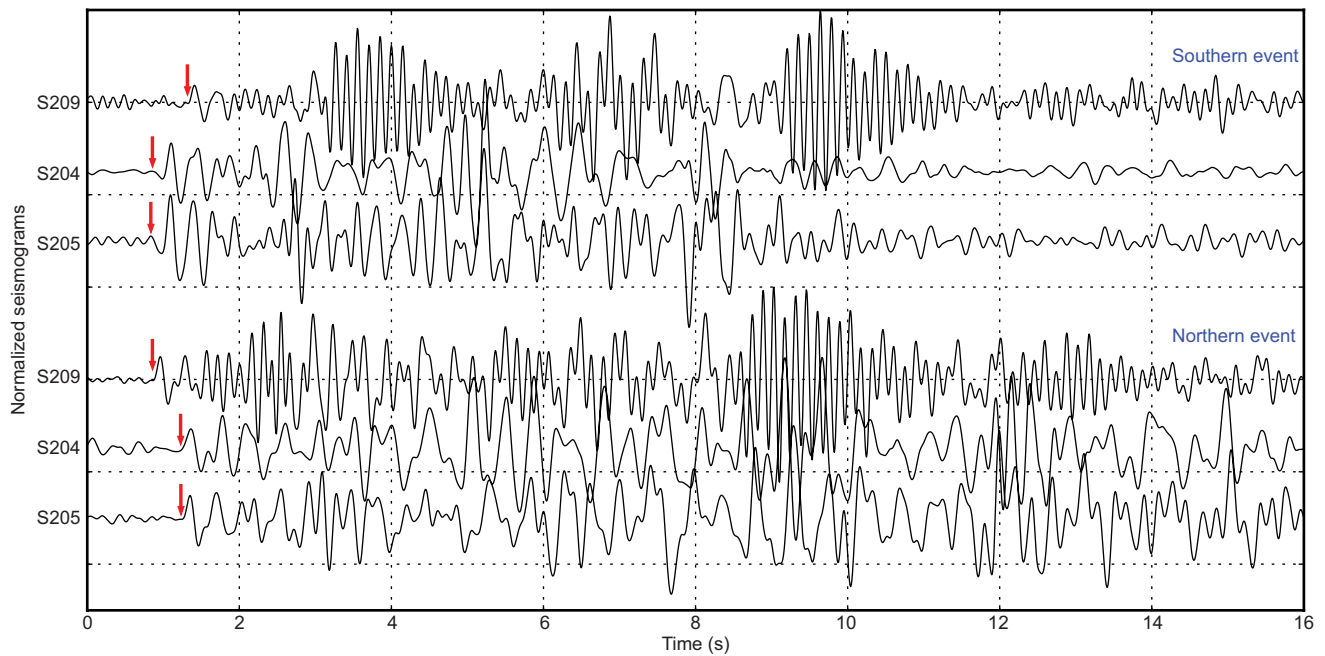


Extended Data Figure 8 | Long-period event locations. **a**, Long-period event locations assuming 0 km depth (black dots) and 1.5 km depth (green stars), assuming a constant crustal velocity of 2.5 km s^{-1} . The red line marks the extent of melt-depleted sub-AML³⁰. Although location depth cannot be definitively determined, the 1.5-h time gap between the initiation of long-period events and melt reaching the surface strongly

supports a deeper source. **b**, Long-period event locations assuming a constant crustal velocity of 2.5 km s^{-1} (green stars) and 2.9 km s^{-1} (black dots) assuming 1.5 km depth. Regardless of the depth and velocity used, events generally cluster near the western edge of the melt lens and bound the area of sub-AML depletion. Faint blue dots represent local earthquake locations (Fig. 1c) for reference.



Extended Data Figure 9 | Long-period event velocity assumption. Minimum of standard deviation for different guess (modelled) velocities, assuming a depth of 1.5 km. The overall minimum of the standard deviation occurs at a velocity of 2.5 km s^{-1} .



Extended Data Figure 10 | Waveforms of long-period events. Waveforms of two separate events at all three stations low-pass filtered at 5 Hz. Red arrows mark the first arrivals. Both event signals arrived at stations S204 and S205 at the same time. This constrains the location of these events

to the west of the ridge axis, between these two stations. The event to the south arrived first at S204 and S205 before S209. The event to the north arrived first at S209 followed by S204 and S205. This constrains the northward migration of the long-period events.

Land-use intensification causes multitrophic homogenization of grassland communities

Martin M. Gossner^{1,2,3}, Thomas M. Lewinsohn^{1,4}, Tiemo Kahl^{5,6}, Fabrice Grassein⁷, Steffen Boch⁷, Daniel Prati⁷, Klaus Birkhofer^{8,9}, Swen C. Renner^{10,11}, Johannes Sikorski¹², Tesfaye Wubet^{13,14}, Hartmut Arndt¹⁵, Vanessa Baumgartner¹², Stefan Blaser⁷, Nico Blüthgen¹⁶, Carmen Börschig¹⁷, Francois Buscot^{13,14}, Tim Diekötter^{18,19}, Leonardo Ré Jorge⁴, Kirsten Jung¹¹, Alexander C. Keyel²⁰, Alexandra-Maria Klein²¹, Sandra Klemmer^{13,22}, Jochen Krauss¹⁷, Markus Lange^{1,2,23}, Jörg Müller²⁴, Jörg Overmann¹², Esther Pašalić^{1,2}, Caterina Penone⁷, David J. Perović^{25,26}, Oliver Purschke^{22,27,28}, Peter Schall²⁹, Stephanie A. Socher³⁰, Ilja Sonnemann³¹, Marco Tschapka¹¹, Teja Tschardt²⁶, Manfred Türke^{1,2,14,22}, Paul Christiaan Venter¹⁵, Christiane N. Weiner¹⁶, Michael Werner¹⁶, Volkmar Wolters¹⁸, Susanne Wurst³¹, Catrin Westphal²⁶, Markus Fischer⁷, Wolfgang W. Weisser^{1,2} & Eric Allan^{7,32}

Land-use intensification is a major driver of biodiversity loss^{1,2}. Alongside reductions in local species diversity, biotic homogenization at larger spatial scales is of great concern for conservation. Biotic homogenization means a decrease in β -diversity (the compositional dissimilarity between sites). Most studies have investigated losses in local (α)-diversity^{1,3} and neglected biodiversity loss at larger spatial scales. Studies addressing β -diversity have focused on single or a few organism groups (for example, ref. 4), and it is thus unknown whether land-use intensification homogenizes communities at different trophic levels, above- and belowground. Here we show that even moderate increases in local land-use intensity (LUI) cause biotic homogenization across microbial, plant and animal groups, both above- and belowground, and that this is largely independent of changes in α -diversity. We analysed a unique grassland biodiversity dataset, with abundances of more than 4,000 species belonging to 12 trophic groups. LUI, and, in particular, high mowing intensity, had consistent effects on β -diversity across groups, causing a homogenization of soil microbial, fungal pathogen, plant and arthropod communities. These effects were nonlinear and the strongest declines in β -diversity occurred in the transition from extensively managed to intermediate intensity grassland. LUI tended to reduce local α -diversity in aboveground groups, whereas the α -diversity increased in belowground groups. Correlations between the β -diversity of different groups, particularly between plants and their consumers, became weaker at high LUI. This suggests a loss of specialist species and is further evidence for biotic homogenization. The consistently negative effects of LUI on landscape-scale

biodiversity underscore the high value of extensively managed grasslands for conserving multitrophic biodiversity and ecosystem service provision. Indeed, biotic homogenization rather than local diversity loss could prove to be the most substantial consequence of land-use intensification.

Land-use intensification threatens biodiversity^{2,4} by reducing the α -diversity of many taxa^{1,3}. Similarly, β -diversity⁵ may decline strongly. This biotic homogenization^{6–9} might occur through either a loss of rare or specialized species (reducing differences between communities), a gain of widespread, generalist species in intensively managed systems (increasing similarity), or most likely a combination of both. Most studies have investigated loss of species richness, but global change may have larger effects on community composition than on local diversity^{10,11}. To separate biotic homogenization from loss of species richness requires measures of β -diversity that distinguish pure species turnover from changes in α -diversity⁵. To predict and manage the loss of β -diversity, we also need to understand whether biotic homogenization occurs at a constant rate as land use intensifies. Land-use intensification can affect α -diversity nonlinearly^{1,12}, but, although environmental gradients can have nonlinear effects on β -diversity^{13,14}, no such effects of land use have been investigated. Here we use data from several landscapes and regions to analyse land-use effects on species turnover (β -turnover)^{15,16} and on total β -diversity (also including differences in species richness) across a wide range of trophic groups.

Different types of organisms probably respond differently to land use. In grasslands, α -diversity belowground may be less affected than aboveground¹. However, land-use intensification may homogenize species composition belowground and reduce β -diversity without reducing

¹Terrestrial Ecology Research Group, Department of Ecology and Ecosystem Management, School of Life Sciences Weihenstephan, Technical University of Munich, Hans-Carl-von-Carlowitz-Platz 2, Freising D-85354, Germany. ²Institute of Ecology, Friedrich-Schiller-University Jena, Dornburger Str 159, Jena D-07743, Germany. ³Swiss Federal Research Institute WSL, Birmensdorf CH-8903, Switzerland. ⁴Department of Animal Biology, IB, UNICAMP—University of Campinas, Campinas, Sao Paulo, CEP, 13083-970, Brazil. ⁵Chair of Silviculture, Faculty of Environment and Natural Resources, University of Freiburg, Tennenbacherstraße 4, Freiburg im Breisgau D-79106, Germany. ⁶Biosphere Reserve Vessertal-Thuringian Forest, Brunnenstr 1, Schmiedefeld am Rennsteig D-98711, Germany. ⁷Institute of Plant Sciences, University of Bern, Altenbergrain 21, Bern CH-3013, Switzerland. ⁸Department of Biology, Biodiversity and Conservation Science, Lund University, Sölvegatan 37, Lund S-22362, Sweden. ⁹Chair of Ecology, Faculty Environment and Natural Sciences, BTU Cottbus-Senftenberg, Großenhainer Str 57, Senftenberg D-01968, Germany. ¹⁰Institute of Zoology, University of Natural Resources and Life Sciences, Wien A-1180, Austria. ¹¹Institute of Evolutionary Ecology and Conservation Genomics, University of Ulm, Ulm D-89069, Germany. ¹²Leibniz-Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Inhoffenstraße 7B, Braunschweig D-38302, Germany. ¹³Department of Soil Ecology, UFZ-Helmholtz Centre for Environmental Research, Halle-Saale D-06120, Germany. ¹⁴Institute of Biology, Leipzig University, Johannisallee 21, Leipzig D-04103, Germany. ¹⁵Biocentre, Institute for Zoology, General Ecology, University of Cologne, Zulpicher Str 47b, Cologne (Köln) D-50674, Germany. ¹⁶Department of Biology, Ecological Networks, Technische Universität Darmstadt, Schnittspahnstraße 3, Darmstadt D-64287, Germany. ¹⁷Department of Animal Ecology and Tropical Biology, Biocentre, University of Würzburg, Am Hubland, Würzburg D-97074, Germany. ¹⁸Animal Ecology, Justus-Liebig-University, Heinrich-Buff-Ring 26-32, Giessen D-35392, Germany. ¹⁹Landscape Ecology, Institute for Natural Resource Conservation, Kiel University, Olshausenstr 75, Kiel D-24118, Germany. ²⁰Department of Ecosystem Modelling, University of Göttingen, Büsgenweg 4, Göttingen D-37077, Germany. ²¹Chair of Nature Conservation and Landscape Ecology, Faculty of Environment and Natural Resources, University of Freiburg, Tennenbacherstraße 4, Freiburg im Breisgau D-79106, Germany. ²²German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, Leipzig D-04103, Germany. ²³Max Planck Institute for Biogeochemistry, Hans-Knoell-Str 10, Jena D-07745, Germany. ²⁴Institute of Biochemistry and Biology, University of Potsdam, Maulbeerallee 1, Potsdam D-14469, Germany. ²⁵Institute of Applied Ecology, Fujian Agriculture and Forestry University, Fuzhou, China. ²⁶Agroecology, Department of Crop Sciences, Georg-August-University Göttingen, Göttingen D-37077, Germany. ²⁷Department of Computer Science, Martin Luther University, Halle-Wittenberg, Halle (Saale) D-06120, Germany. ²⁸Geobotany and Botanical Garden, Institute of Biology, Martin Luther University, Halle-Wittenberg, Halle (Saale) D-06108, Germany. ²⁹Department Silviculture and Forest Ecology of the Temperate Zones, University of Göttingen, Göttingen D-37077, Germany. ³⁰Department of Ecology and Evolution, Botanical Garden, University of Salzburg, Hellbrunnerstrasse 34, Salzburg 5020, Austria. ³¹Functional Biodiversity, Institute of Biology, Freie Universität Berlin, Königin-Luise-Str. 1-3, Berlin D-14195, Germany. ³²Centre for Development and Environment, University of Bern, Hallerstrasse, 10, Bern CH-3012, Switzerland.

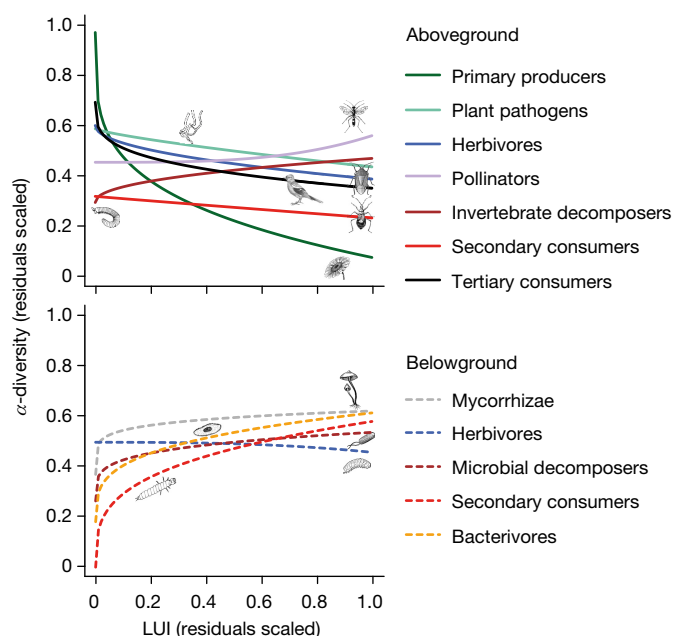


Figure 1 | The effect of LUI on α -diversity above- and belowground.

The partial effect of local LUI was generated with a power law model fitted to the species richness of the seven aboveground (solid lines) and the five belowground trophic groups (dashed lines) ($n = 105$ plots; for details, see Methods). Species richness and LUI were corrected for differences due to region, pH, soil nutrients, sdLUI (standard deviation of LUI across five years) and isolation, by taking residuals, and were then scaled between 0 and 1.

α -diversity, as shown for tropical soil bacteria¹⁷. Biotic homogenization might therefore be widespread, but at present we lack a broader view of how land-use intensification alters α - and β -diversity of distinct functional or trophic groups. The loss of specialist species, which generally accompanies biotic homogenization, could also reduce correlations between the β -diversities of different groups¹⁸, indicating changes in trophic structure.

We compiled a unique set of biodiversity data, including more than 4,000 species of plants, arthropods, foliar fungal pathogens, mycorrhizal fungi, bacteria, protists, bats and birds, measured on 105 grasslands varying in local LUI (a compound index of grazing, mowing and fertilization intensity)¹⁹. We divided the species into 12 groups on the basis of trophic level and whether they lived above- or belowground. We modelled the effect of LUI on α - and β -diversity of each group, accounting for potential environmental and spatial effects. We assessed biotic homogenization in three ways, by testing: (1) for an overall negative effect of land use on β -diversity; (2) where along the land-use gradient the greatest change in β -diversity occurred; and (3) whether correlations among β -diversity of different trophic groups were reduced, which would indicate a loss of specialist species.

High LUI reduced α -diversity for most aboveground groups, but had neutral or positive effects on belowground organisms (Fig. 1). These results were consistent regardless of the weight given to common species, that is, whether α -diversity was measured as species richness, Shannon or Simpson diversity (Extended Data Fig. 1). Land-use effects were not driven by co-varying environmental factors because we adjusted for soil pH, nutrients and geography (see Methods). Differences between above- and belowground communities may occur because they respond at different spatial scales²⁰ and belowground groups are better protected from disturbance²¹. Alternatively, a shift towards bacterial-dominated communities in more intensively managed grasslands²² may have cascaded up to increase the diversity of higher trophic levels. Nevertheless, these opposing above- and belowground responses were unanticipated and have not previously been shown in a multitrophic dataset.

We then analysed the effects of land-use on β -diversity and found widespread evidence for biotic homogenization both above and below ground. We modelled β -diversity between all possible plot pairs using linear models, again correcting for other environmental and geographic drivers, and including two descriptors of LUI: the mean and the difference in LUI (Δ LUI) between them. The mean LUI represents overall intensity; any negative effects on β -diversity indicate biotic homogenization. The linear Δ LUI term represents land-use heterogeneity and positive effects indicate that mixing grasslands of low and high LUI increases β -diversity. Increasing land-use intensity (mean LUI) had strong negative effects on the β -turnover of many above- (4 out of 7) and belowground (2 out of 5) groups (Fig. 2a, Extended Data Figs 2 and 3), indicating biotic homogenization both above- and belowground, in contrast to the opposing responses of α -diversity. Belowground groups, especially mycorrhizae and bacterivores, were therefore homogenized at high LUI even though their α -diversity increased. These stronger effects of LUI on belowground β -turnover extend findings from Amazonian bacterial communities, responding to marked changes in land use¹⁷, to a much larger number of groups. For many groups, increasing LUI had an even larger effect on total β -diversity, which includes changes in species richness and turnover (Extended Data Fig. 4). This was particularly evident for plants, which suffered substantial species loss. In general, Δ LUI effects were smaller than mean LUI effects, showing that increasing land-use heterogeneity has limited potential to offset negative effects of intensification. Land-use heterogeneity might be even less beneficial in cases where high LUI grasslands are dominated by exotic species of low conservation value. Our multitrophic results suggest that, despite their differences in dispersal rates and body sizes^{23,24}, large-scale spatial dynamics are similar in below- and aboveground groups.

We next investigated whether the rate of biotic homogenization was constant over the LUI gradient, and found that it peaked in the transition from low to intermediate LUI. Using generalized dissimilarity modelling (GDM) we fitted nonlinear effects of Δ LUI on β -diversity along the LUI gradient¹³. LUI was a major driver of β -turnover and total β -diversity, even compared to the large spatial and nutrient differences between the grasslands (deviance explained in Fig. 2b and relative effects in Extended Data Figs 5–7; Supplementary Information Section 5). There was a general trend for saturating responses in the β -diversity of aboveground (plants, herbivores and pollinators) and belowground groups (bacterivores and mycorrhizae) (Fig. 2b), which was parallel to the α -diversity response aboveground¹. Differences in LUI between grasslands therefore drive turnover only at low overall LUI and increasing land-use heterogeneity beyond a certain point will not increase β -diversity, supporting the conclusion that minimizing LUI across the landscape most effectively enhances β -diversity. Some other groups, plant pathogens and secondary consumers, showed accelerating responses in β -diversity, which indicates strong homogenization in the most intensively managed grasslands. When we analysed the effects of the LUI components (grazing, mowing and fertilization) separately in the GDMs, mowing intensity was the main driver of biotic homogenization for most groups (Extended Data Fig. 8). Frequent mowing creates a homogenous sward, reduces flowering and seed set, causes high insect mortality and may lead to soil compaction, all of which may cause extinctions of rare species and favour a small set of disturbance-tolerant species both above- and belowground. In a global analysis, elevated nutrient input proved to be a main driver of soil microbial community composition²⁵. In our study, fertilization had comparatively minor effects: increased homogeneity in soil nutrient levels at high LUI seemed to reduce β -diversity (see Supplementary Table 5 for LUI results without soil nutrients) less than homogenization of disturbance regimes.

Effects of LUI on total β -diversity were generally larger when measurements were weighted by species abundances (Extended Data Fig. 4). Intensively managed grasslands may be dominated by the same common species, even if they differ in their rare species. Indeed, a common

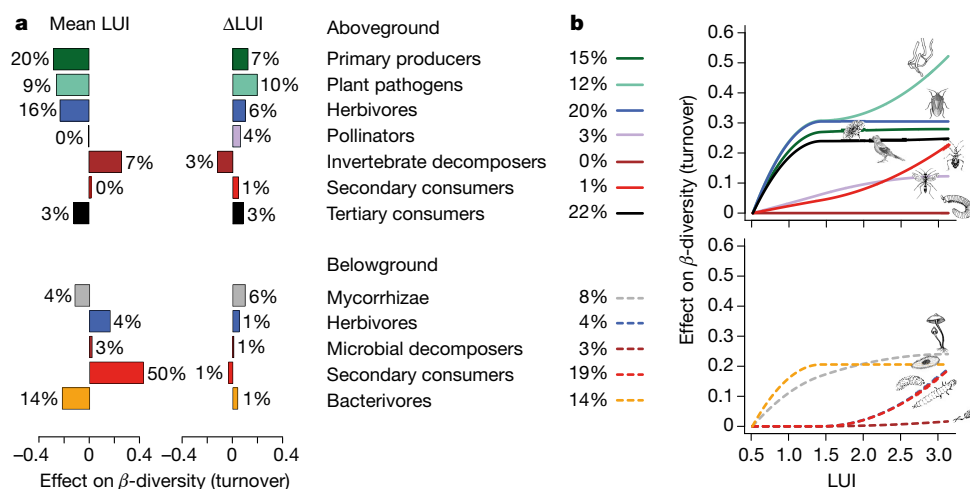


Figure 2 | Effects of LUI on β -diversity above- and belowground.

a, Bars show partial effects of mean LUI and differences in LUI (Δ LUI) between plot pairs (105 plots), on species turnover (β_{sim}), from linear models. Numbers adjoining bars are the proportion of explained variance uniquely explained by mean LUI or Δ LUI. **b**, Results from the generalized

dissimilarity models (GDMs) showing the effect of Δ LUI on β_{sim} along the LUI gradient, with percentages of deviance uniquely explained by LUI. Higher maximum curves indicate larger effects. All effects are corrected for environmental covariates and explanatory variables are scaled to allow comparisons across trophic levels.

practice in grassland management is to seed intensively used plots with a few fast-growing species of high nutritional value, which reinforces the homogenization of plant communities under high LUI. Increased abundance of common, generalist species might also drive biotic homogenization in other trophic groups. Under high LUI, dominance increased in most aboveground groups, although not in belowground groups (Extended Data Fig. 1). Increased dominance by a small set of common species, across a wide array of trophic groups, might threaten the delivery of critical ecosystem services in intensively managed landscapes²⁶.

Despite the overall consistency of land-use effects, some exceptions are worth noting. Bacteria responded weakly and had very low β -diversity, perhaps because their taxonomic resolution was coarser than for other groups (Methods). Responses of pollinators were also weak,

possibly because their β -diversity responds more to land use at the landscape scale²⁷, as shown by the strong response to grassland isolation (Extended Data Fig. 5, Supplementary Information Section 5). In most other groups, isolation was a much less important driver of β -diversity. Only in three invertebrate groups did β -diversity increase with LUI; however, these groups were species-poor and had a lower sample coverage (see Supplementary Information Section 2 and Extended Data Fig. 9). The relative importance of LUI as a driver of β -diversity therefore varied between trophic groups (Extended Data Fig. 7), but it affected key ecosystem service providers such as plants and herbivores, as well as rare birds, which have a high conservation value.

High LUI also homogenized trophic structure and disrupted correlations between β -diversity of adjacent trophic levels. We calculated correlations between β -diversities for sets of plots with low, versus high,

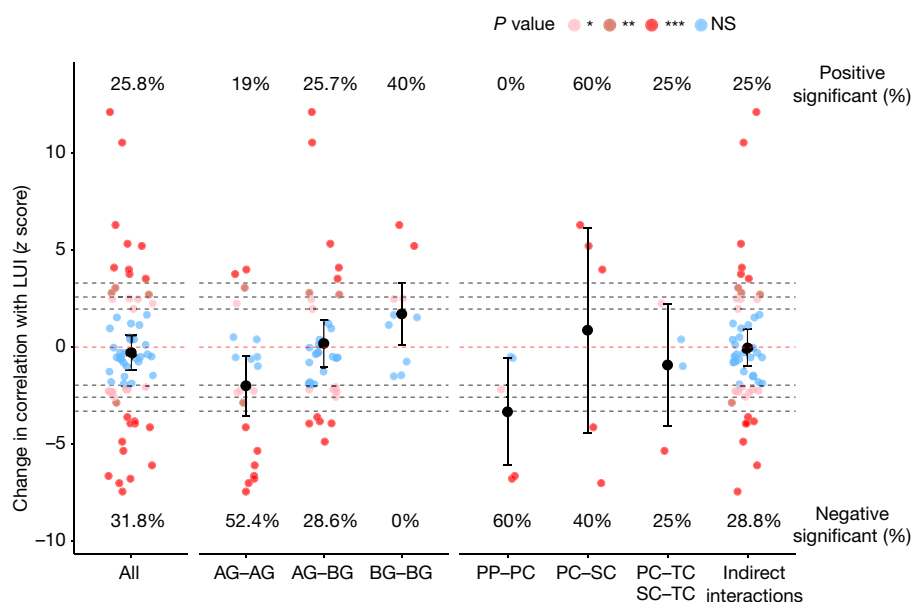


Figure 3 | Effect of LUI on correlations between the β -diversities (β_{sim}) of different trophic groups. z scores (positive z scores indicate that correlations are higher at high LUI, and negative z scores indicate that correlations are lower at high LUI) and P values (dashed lines separate P levels; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, NS, not significant) show the change in correlation strength (R^2 values from matrix regressions, corrected for effects of differences in LUI) between low (52 plots) and high

LUI (53), comparing observed values to random values (see Methods). All correlations were grouped into categories: AG, aboveground; BG, belowground; PP, primary producers; PC, primary consumers; SC, secondary consumers; TC, tertiary consumers. Each coloured dot represents one correlation, black dots represent the mean and black bars the 95% confidence intervals. For statistical details see Supplementary Information Section 5.

LUI and expected a drop in correlations at high LUI, which would indicate biotic homogenization. Correlation strength dropped by more than 50% on average (for $R^2 > 0.1$ at low LUI) at high LUI and correlations between aboveground groups and between producers and primary consumers (plants and herbivores or pathogens) declined substantially (Fig. 3 and Extended Data Fig. 10), potentially reflecting a loss of host specialists. Some correlations increased in strength but these mainly involved the species-poor invertebrate groups whose β -diversity increased with LUI (see Supplementary Table 6). A previous study²⁸ showed that land-use intensification disrupted correlations in α -diversity and we extend this finding to show spatial decoupling for a wider range of trophic groups.

By analysing a uniquely comprehensive biodiversity dataset, we showed that LUI substantially reduces β -diversity across many different trophic groups. This threatens biodiversity by homogenizing communities within and across agricultural landscapes. The consequences of biotic homogenization for landscape-scale ecosystem service provisioning remain uncertain, but are likely to be severe²⁶. Moreover, our results show that measures to reduce management intensity should be most effective at intermediate LUI¹²; they also underscore the high value of extensively managed grassland for conserving multitrophic diversity by showing that reduced intensity across the landscape effectively promotes large-scale diversity. This could be achieved by increasing the area of extensively managed grasslands in general, and especially by reducing the intensity of mowing. Conservation strategies and agricultural policies will increase in effectiveness if they aim to maintain the heterogeneity of biotic communities at the landscape scale, for instance by coupling subsidies to landscape-scale measures of diversity.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 June; accepted 25 October 2016.

Published online 30 November 2016.

- Allan, E. *et al.* Interannual variation in land-use intensity enhances grassland multidiversity. *Proc. Natl Acad. Sci. USA* **111**, 308–313 (2014).
- Sala, O. E. *et al.* Global biodiversity scenarios for the year 2100. *Science* **287**, 1770–1774 (2000).
- Newbold, T. *et al.* Global effects of land use on local terrestrial biodiversity. *Nature* **520**, 45–50 (2015).
- Foley, J. A. *et al.* Global consequences of land use. *Science* **309**, 570–574 (2005).
- Anderson, M. J. *et al.* Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecol. Lett.* **14**, 19–28 (2011).
- Gámez-Virués, S. *et al.* Landscape simplification filters species traits and drives biotic homogenization. *Nat. Commun.* **6**, 8568 (2015).
- Karp, D. S. *et al.* Intensive agriculture erodes β -diversity at large scales. *Ecol. Lett.* **15**, 963–970 (2012).
- McKinney, M. L. & Lockwood, J. L. Biotic homogenization: a few winners replacing many losers in the next mass extinction. *Trends Ecol. Evol.* **14**, 450–453 (1999).
- Smart, S. M. *et al.* Biotic homogenization and changes in species diversity across human-modified ecosystems. *Proc. R. Soc. B* **273**, 2659–2665 (2006).
- Dornelas, M. *et al.* Assemblage time series reveal biodiversity change but not systematic loss. *Science* **344**, 296–299 (2014).
- Solar, R. R. C. *et al.* How pervasive is biotic homogenization in human-modified tropical forest landscapes? *Ecol. Lett.* **18**, 1108–1118 (2015).
- Kleijn, D. *et al.* On the relationship between farmland biodiversity and land-use intensity in Europe. *Proc. R. Soc. B* **276**, 903–909 (2009).
- Ferrier, S., Manion, G., Elith, J. & Richardson, K. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers. Distrib.* **13**, 252–264 (2007).
- Fitzpatrick, M. C. *et al.* Environmental and historical imprints on beta diversity: insights from variation in rates of species turnover along gradients. *Proc. R. Soc. B* **280**, 20131201 (2013).
- Carvalho, J. C., Cardoso, P., Borges, P. A. V., Schmera, D. & Podani, J. Measuring fractions of beta diversity and their relationships to nestedness: a theoretical and empirical comparison of novel approaches. *Oikos* **122**, 825–834 (2013).
- Baselga, A. & Leprieux, F. Comparing methods to separate components of beta diversity. *Methods Ecol. Evol.* **6**, 1069–1079 (2015).
- Rodrigues, J. L. M. *et al.* Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. *Proc. Natl Acad. Sci. USA* **110**, 988–993 (2013).
- Pellissier, L. *et al.* Turnover of plant lineages shapes herbivore phylogenetic beta diversity along ecological gradients. *Ecol. Lett.* **16**, 600–608 (2013).
- Blüthgen, N. *et al.* A quantitative index of land-use intensity in grasslands: integrating mowing, grazing and fertilization. *Basic Appl. Ecol.* **13**, 207–220 (2012).
- De Deyn, G. B. & Van der Putten, W. H. Linking aboveground and belowground diversity. *Trends Ecol. Evol.* **20**, 625–633 (2005).
- Haimi, J., Fritze, H. & Moilanen, P. Responses of soil decomposer animals to wood-ash fertilisation and burning in a coniferous forest stand. *For. Ecol. Manage.* **129**, 53–61 (2000).
- Bardgett, R., Hobbs, P. & Frostegård, Å. Changes in soil fungal:bacterial biomass ratios following reductions in the intensity of management of an upland grassland. *Biol. Fertil. Soils* **22**, 261–264 (1996).
- Fenchel, T. & Finlay, B. J. The ubiquity of small species: patterns of local and global diversity. *Bioscience* **54**, 777–784 (2004).
- Finlay, B. J. Global dispersal of free-living microbial eukaryote species. *Science* **296**, 1061–1063 (2002).
- Leff, J. W. *et al.* Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proc. Natl Acad. Sci. USA* **112**, 10967–10972 (2015).
- van der Plas, F. *et al.* Biotic homogenization can decrease landscape-scale forest multifunctionality. *Proc. Natl Acad. Sci. USA* **113**, 3557–3562 (2016).
- Perović, D. *et al.* Configurational landscape heterogeneity shapes functional community composition of grassland butterflies. *J. Appl. Ecol.* **52**, 505–513 (2015).
- Manning, P. *et al.* Grassland management intensification weakens the associations among the diversities of multiple plant and animal taxa. *Ecology* **96**, 1492–1501 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to J. Chase and M. Fitzpatrick for their comments and suggestions on a previous version of the manuscript; B. Büche, R. Achtziger, T. Wagner, F. Köhler, T. Blick and M.-A. Fritze for arthropod species identification and U. Kern for creating the small icons of the 12 trophic groups used in the figures. We thank the managers of the three Exploratories, K. Hartwich, S. Gockel, K. Wiesner and M. Gorke for their work in maintaining the plot and project infrastructure; C. Fischer and S. Pfeiffer for giving support through the central office, M. Owonibi for managing the central data base; and E. Linsenmair, D. Hennenmöller, J. Nieschulze, I. Schöning and the late E. Kalko for their role in setting up the Biodiversity Exploratories project. We are also grateful to E. Kalko for her invaluable inspiration and for launching the studies on bats and birds. The work has been funded by the DFG Priority Program 1374 'Infrastructure-Biodiversity-Exploratories'. Field work permits were issued by the responsible state environmental offices of Baden-Württemberg, Thüringen and Brandenburg (according to §72 BbgNatSchG).

Author Contributions M.M.G. and E.A. conceived the idea for the manuscript, and defined the final analysis. M.M.G., E.A., C.P., T.M.L. and T.K. analysed the data. M.M.G. and E.A. wrote the first manuscript draft and finalized the manuscript. A.M.K., C.B., C.N.W., C.W., D.J.P., D.P., E.P., F.B., H.A., I.S., J.K., J.M., J.S., J.O., K.J., K.B., M.T., M.F., M.L., M.M.G., M.W., N.B., P.C.V., S.B.I., S.Bo., S.A.S., S.C.R., S.K., S.W., T.D., T.W., V.B., V.W., and W.W.W. contributed data. T.M.L., F.G., S.Bo., D.P., L.R.J., K.B., S.C.R., A.C.K., O.P., P.S., T.T., W.W.W. and J.S. contributed substantially to revisions. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M.G. (martin.gossner@wsl.ch).

Reviewer Information Nature thanks P. Barton, S. Prober and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

Study system. The study was conducted as part of the Biodiversity Exploratories project (www.biodiversity-exploratories.de) in three German regions: (1) The UNESCO Biosphere area Schwäbische Alb in the low mountain range in south-western Germany (420 km², 460–860 m above sea level (a.s.l.)); (2) the Hainich National Park and its surrounding areas in the hilly lands of central Germany (1560 km², 285–550 m a.s.l.); and (3) the UNESCO Biosphere Reserve Schorfheide-Chorin in the glacially formed lowlands of northeastern Germany (1300 km² in size, 3–140 m a.s.l.). The three regions differ in climate, geology and topography, but each is characterized by a gradient of LUIs typical for large parts of temperate Europe²⁹. Land-use gradients in this study range from semi-natural to intensively managed grasslands, because natural grasslands, that do not require management to prevent succession to forest, are almost absent from western and central Europe. In each region, 50 grassland plots were chosen from a total of 500 candidate plots, on which initial vegetation and land-use surveys were conducted, by stratified random sampling. This ensured that the plots covered the whole range of LUIs and management types, while minimizing confounding factors such as spatial position or soil type²⁹. Thereby we avoided, for instance, sampling low-intensity grasslands only in the low productive parts of the landscape. All grasslands have a long history of broadly similar LUI (that is, low intensity grasslands have not been recently converted from high intensity grasslands and vice versa and all had been grasslands for at least 10 years before the start of project), although we are aware that temporal variation in land use is substantial¹. In this study, we analysed a subset of 105 plots (Schwäbische Alb: 32, Hainich-Dün: 37, Schorfheide-Chorin: 36) for which data on all taxa (see below) were available.

Study design and land-use measures. All plots were continuously managed by farmers. Information on management practices, including the level of fertilization (kg N ha⁻¹ yr⁻¹), grazing (number of livestock units ha⁻¹ yr⁻¹) and mowing (number of cuts per year), was assessed annually by standardized interviews with the land owners. LUI at the local scale was then quantified as a compound index on the basis of summing the standardized intensities of these three components¹⁹. We decided to employ a compound index of LUI because the individual components are correlated with each other (fertilization and mowing are positively correlated, and grazing and mowing negatively correlated¹⁹) and the distribution of overall intensity is more even (each individual component has many 0 values). Each component was divided by the global mean value for each year to standardize the components¹⁹. We then calculated the mean LUI for each plot over five years (2006–2010) because this reflects the average LUI around the years when most of the data was assessed (2008, 2009 and 2011). At the low end of land-use intensity, with a LUI of 0.5, grasslands are typically unfertilized, not mown, and grazed by 40–50 sheep per hectare for about 10 days (or more rarely by 1–3 cattle per hectare for 20 days). At an intermediate LUI of 1.5, grasslands are usually unfertilized (or fertilized with less than 30 kg N ha⁻¹ yr⁻¹), and are either mown twice a year or grazed by four cattle per hectare for about 50 days. At the high end of land-use intensity with a LUI of 3, grasslands are typically fertilized at a rate of 60–120 kg N ha⁻¹ yr⁻¹, are mown 2–3 times a year or grazed by 5–10 cattle for 100–150 days, or are managed by a combination of grazing and mowing. In addition to using the LUI index (that is, where all three types of land use are given equal weight), we tested the individual standardized land-use components in our models to separate nutrient and disturbance effects.

Covariates. We corrected for a series of other variables that might affect diversity. In addition to the mean land use across time, we quantified two other measures of land use: (1) the temporal variation of LUI, as its standard deviation (sdLUI) across five years (2006–2010) because this has been shown to increase α -diversity¹; and (2) to measure effects of land use at the landscape scale, which may also be important drivers of diversity⁶, we calculated one minus the proportion of grasslands (including managed grasslands and semi-natural vegetation and thus all potential habitat for grassland species) within 500 m of each plot as a proxy for isolation. Plot neighbourhood cover was mapped in 2009 using high resolution (40 cm) aerial photographs from 2008, including the following land-cover types besides grasslands: arable land, forest, roads, trees (woodlots smaller than 1 ha), urban areas and water bodies.

We also corrected for soil nutrients and pH. We sampled the upper 10 cm of the mineral soil to assess soil variables. This was conducted in May 2011 with 14 locations per plot along two 18-m transects with distances of 3 m between sampling points, for details see ref. 30 and Supplementary Information Section 1. Soil nutrients were quantified by the first axis of a principal component analysis (PCA, ade4 package³¹) of nitrogen stock, total soil nitrogen, organic carbon stock, soil inorganic carbon, soil organic carbon, soil C:N ratio (Supplementary Table 1).

Note that our measure of soil nutrients partly includes, in addition to natural variation among soils, the effect of fertilization, which in turn is also part of the LUI. Therefore, our estimates on the effects of LUI are on the conservative side and fitting models without the soil nutrient variable slightly increases the strength of the LUI effects in some cases, see Supplementary Information Sections 4, 5. Finally, in the analysis of β -diversity, we also corrected for spatial distance between all sites. To consider spatial effects on β -diversity, we calculated geographic distance between each pair of plots on the basis of geographic coordinates using Euclidean distances in the R package *vegan*³².

Biodiversity assessment. We assessed biodiversity across a broad range of organisms from bacteria to vertebrates using molecular (bacteria, arbuscular mycorrhizal fungi and protists) as well as morphological (plants, arthropods and pathogenic fungi) or acoustic characteristics (bird and bat calls).

Vascular plants, bryophytes and lichens. We sampled vascular plants between mid-May and mid-June 2009, bryophytes (2007–2008) and lichens (2007–2008) in an area of 4 m × 4 m on each plot, and estimated the percentage cover of each occurring species. For details see refs 33, 34 for vascular plants, ref. 35 for lichens and ref. 36 for bryophytes).

Arthropods. All arthropods of the herb layer were sampled in 2008 by different methods. For sampling Araneae, Coleoptera, Hemiptera: Heteroptera and Auchenorrhyncha, Hymenoptera: Symphyta, Neuroptera, Orthoptera, Dermaptera and Dictyoptera we used biannual (June and August) sweep netting by conducting 60 double sweeps along three 50-m long plot border transects^{37,38}. Additionally, Diptera and Hymenoptera were hand-collected during their visits on flowers, identified and individuals counted³⁹. This survey involved a transect of 200 m × 3 m along the edge of the plot, for which three transect walks were performed on a single day (total, 6 h). In some cases, plots were measured several times; these were averaged in less than one month apart or, if repeated over one month later, the earlier measure was used³⁹. We conducted surveys of butterflies and day-active moths (hereafter termed as Lepidoptera) from beginning of May to mid-August^{40,41}. We sampled Lepidoptera on fixed transects in the three regions repeating the sampling three times in a randomized sequence within each region. Each transect had a length of 300 m and we recorded all Lepidoptera within 30 min per site within a 5 m corridor.

Soil arthropods (Myriapoda, soil living larvae) were sampled in spring 2011 (within ten days in April) by collecting two soil cores (diameter, 20 cm, depth, 10 cm) from each plot. Soil fauna was extracted from the first core using a modified heat extraction system⁴² over a period of eight days and the second soil core was hand-sorted for soil macrofauna.

All arthropod species were assigned to one of four trophic groups (herbivores, pollinators, predators and decomposers) on the basis of their known main food resource as adults.

Pathogenic fungi. From July to August 2011, we sampled pathogenic fungi including rust, powdery mildew, downy mildew and smut fungi in four transects of 25 m × 1 m per plot. We inspected all occurring vascular plant species for infested individuals, sampled them and later identified the pathogenic fungi to the species level.

Birds. Birds were sampled by audio-visual point counts⁴³ covering the area of the respective grassland plot (50 m × 50 m). We noted all individuals of each bird species during the five-minute interval. In each year, from 2008 to 2012, we visited each plot five times between 15 March and 15 June (1st surveying period, 15–30 March; 2nd, 15–30 April; 3rd, 1–15 May; 4th, 16–31 May; 5th, 1–15 June). A maximum of 15 plots was surveyed per day from sunrise to 11:00; occasionally the evening chorus was surveyed after 17:00 (<20 times out of 750 events per year). The sequence in which plots were visited was randomized. The maximum number of birds displaying per site per year (that is, the maximum number of record individuals per species over the five rounds within a surveying year) was used as a measure of the relative abundance of birds. We considered a species as present in the particular plot if it was recorded at least once during a survey within each year. Aerial species (swifts and swallows) were excluded from analysis, as they had only been surveyed irregularly and without standardization. For this study, we combined species richness and relative abundance data across the five sampling periods.

Bats. We surveyed bats from June to September during the years 2008, 2009 and 2010. Plot sampling was conducted along a 24 min point-stop transect of 200 m at the borders of each grassland plot. Sampling started 30 min after local sunset and was limited to the first half of the night (01:00) to account for the first peak in bat activity⁴⁴. We randomly sampled 4–6 plots per night. Each plot was surveyed twice during each year with a minimum time interval between repeated sampling of five weeks. Acoustic recordings of bats were taken in real time (sample rate: 384 kHz, 16 bit) with a Pettersson-D1000x bat detector (Pettersson Electronic AG, Uppsala, Sweden) and triggered manually by an observer listening through headphones to the output of the heterodyne system while continuously scanning the frequency

range between 20 and 80 kHz. Bat species identification was conducted using Avisoft SAS Laboratory Pro, Version 5.0.24 (R. Specht, Avisoft Bioacoustics, Berlin, Germany, Hamming window, 1024 FFT, 96% overlap) following various references on echolocation call parameters, for example refs 45–47. For details on species identification see ref. 48. In addition, we evaluated the number of bat passes that were defined as a minimum of two consecutive echolocation calls⁴⁹. Successive passes within one recording were discriminated if the time interval between calls was larger than three times the regular pulse interval of the particular species^{48,50}.

Belowground microorganisms. At each grassland plot fourteen soil cores (diameter, 8.3 cm) were taken from a 20 m × 20 m subarea and soil from the upper 10 cm of the A horizon was homogenized after removal of root material. The bulk sample was split into subsamples for the analyses of bacteria, protists and arbuscular mycorrhizal fungi.

For bacteria, 10 g of the homogenized soil was put immediately on liquid nitrogen and stored until RNA extraction⁵¹. Briefly, total RNA was isolated from soils and reverse transcribed into cDNA. Amplicons of the V3 region of the 16S rRNA gene were sequenced on an Illumina HiSeq platform using universal bacterial primers as described in ref. 52.

For the analysis of protists, 1 g of the bulk soil sample was used for DNA extraction and the analyses of the V4 region of the 18S rRNA gene amplified using eukaryotic specific primers. Sequences were filtered for (1) 100% forward primer match; (2) length ≥ 200–710 bp and (3) ambiguities (N). Traces were scanned for chimaeras, trimmed to 530 bp, dereplicated to group 100% identical amplicons, and singletons removed. Remaining sequences were treated as operational taxonomic units (OTUs) and aligned to the PR2 database using BLASTn (default parameters). One hit per sequence was retained. Only OTUs with 100% coverage and protist taxa (excluding Metazoa, Fungi and Streptophyta) were retained for analysis.

For the study of arbuscular mycorrhizal fungi, total microbial DNA was isolated from the bulk soil sample using a MoBioPowerSoil DNA Isolation Kit. The NS31-a.m.1 fragment of the fungal 18S rDNA was amplified using arbuscular mycorrhizal fungal specific primers⁵³ and sequenced using a Genome Sequencer FLX+ 454 System. The reads were quality filtered using Mothur⁵⁴ and classified using the MaarjAM AMF reference database⁵⁵. A total of 825 arbuscular mycorrhizal fungal OTUs were detected.

Detailed description of the data processing of bacteria, protists and arbuscular mycorrhizal fungi is presented in Supplementary Information Section 2.

Statistical analyses. All analyses were conducted in R 3.0.2 (ref. 56).

Sample completeness. To test for sample completeness, we used a previously published approach of sample coverage^{57,58}. Coverage is defined as the proportion of the total number of individuals in an assemblage that belong to species represented in the sample. We used two approaches to estimate sample coverage. First, we estimated sample coverage for low (52) and high (53) LUI plots on the basis of species incidences. Second, we estimated sample coverage for each plot on the basis of species abundances. Sample coverage did not differ significantly along the LUI gradient and was estimated to be higher than 90% in all trophic groups, except aboveground invertebrate decomposers and secondary consumers (Extended Data Fig. 10 and Supplementary Table 2-2). Therefore, the results for secondary consumers and invertebrate decomposers should be treated with caution but there is no evidence that any undersampling is biased along the LUI gradient. A further line of evidence for the robustness of our findings to issues of sample completeness is provided by the results for different q -levels. When increasing q -levels, rare species are less strongly weighted in the calculation of β -diversity. As we found similar or even stronger responses of β -diversity to LUI at higher q -levels (see below) this further indicates that undersampling of rare species is unlikely to affect the conclusions of our study. Analyses were conducted using the iNEXT function in the iNEXT library⁵⁹.

Diversity measures. We calculated several measures of α - and β -diversity for each of the 12 trophic groups (Fig. 1). We used previously published q -metrics^{60,61} to incorporate different weightings for species abundance in α - and β -diversity. These are based on Hill numbers⁶², and allow the calculation of diversity measures in which increasing weight is given to species abundances. At $q = 0$, rare and abundant species are weighted equally, which corresponds to species richness for α -diversity and the Sørensen index of dissimilarity for β -diversity. At $q = 1$, species are weighted in proportion to their frequency in the sampled community, which corresponds to the exponential of Shannon entropy (or effective number of species) for α -diversity and Horn's index of dissimilarity for β -diversity. Finally, at $q = 2$, abundant species receive more weight relative to their frequency and this corresponds to the inverse Simpson index for β -diversity and Morisita–Horn index for β -diversity^{63,64}. Analyses were conducted using multipart function in the vegan library⁶⁵.

Spatial turnover in composition between locations involves two main processes: a replacement of species (pure turnover) and changes in species richness^{15,16}. To test

for effects on β -diversity that are independent of species richness differences, we used the Simpson dissimilarity β_{sim} , which is the turnover component of Sørensen dissimilarity, see also ref. 16. Details on β - (diversity across the three study regions) as well as β -diversity partitioning is given in Supplementary Information Section 3.

α -diversity analysis. To analyse the response of α -diversity to land use, we used power law models that allow different shapes of responses to be fitted. We modelled the response of α -diversity for each of the 12 trophic groups, calculated with $q = 0, 1$ or 2 . The explanatory variable was LUI, the model formula was $y = a + (b \times LUI^c)$, where a (intercept), b (slope) and c (degree of curvature) are parameters estimated by the model. In order to correct for confounding environmental effects we analysed residuals in the power law models. We calculated residuals from linear models with diversity or land use (LUI) as the response variable and region, soil nutrients, pH, variation in LUI (sdLUI) and isolation (1 – proportion of grasslands in the plot surrounding) as explanatory variables. We calculated residuals because incorporating many explanatory variables in the power law models would have led to extremely complex models. After taking residuals, we then scaled all explanatory and response variables between 0 and 1, to allow comparison of effects and responses. Models were fitted using the gnl function in the nlme library⁶⁶.

β -diversity analysis. Linear models. To analyse the response of β -diversity to land-use intensification, we first fitted linear models. These were fitted to values of turnover (β_{sim}), total β -diversity (Sørensen, $q = 0$) and abundance weighted β -diversity ($q = 1$ or 2) for each of the 12 trophic groups. Explanatory variables in these models were: the mean LUI, sdLUI, isolation, soil nutrients and pH between each pair of plots. The effects of mean LUI provide a test of biotic homogenization: a negative effect indicates that land-use intensification reduces turnover. In this case, reducing LUI across the landscape would promote β -diversity. However, as plot pairs with the same mean LUI can either be very similar in LUI or come from different ends of the gradient, we also fitted differences in LUI (ΔLUI) between all plot pairs. This term tests for the effect of land-use heterogeneity, that is, whether β -diversity is higher between plots of different intensities. A positive effect would suggest that maximizing land-use heterogeneity across the landscape would increase β -diversity. The terms mean LUI and ΔLUI are not correlated with each other although ΔLUI is constrained to zero at maximum and minimum mean LUI. We additionally fitted differences in all other variables (sdLUI, isolation, soil nutrients and pH) together with the spatial distance between all plot pairs in the models. To compare the effects of the different predictors, we scaled all predictors to between 0 and 1. We then calculated the variance explained uniquely by mean LUI or ΔLUI by comparing the variance explained by the full model with that explained by models containing all terms except mean LUI or ΔLUI . The unique variance is expressed as a proportion of the total explained variance. We calculated the significance of all terms in the linear models using a permutation procedure, implemented with the lmp function in the lmp library⁶⁷, using 100,000 iterations.

We also ran a second series of linear models in which we replaced mean LUI with the mean grazing, mowing and fertilization intensity between plot pairs (all scaled to the maximum across plots) and replaced ΔLUI with differences in grazing, mowing and fertilization between plots. Finally, we ran linear models without soil nutrient levels, to test whether effects of LUI were mediated by its effects on soil nutrients.

Generalized dissimilarity modelling. To analyse the nonlinear effects of differences in LUI, we used GDM¹³. This is a matrix regression technique for modelling turnover in species composition between sites as a function of the spatial and/or environmental distance between them. The advantages of GDM are that it can incorporate variation in the rate of compositional turnover along an environmental or spatial gradient (non-stationarity) and that it allows the relationships between dissimilarity and distance to be nonlinear. The only constraint is that compositional turnover is assumed to always increase with distance between sites (monotonicity). For more details on GDM see ref. 13. All GDMs were fitted using the gdm function in the gdm library⁶⁸. We plot the partial effect of each predictor, that is, while holding all other predictors constant, against the level of a given predictors to visualize the results of the GDM (Extended Data Fig. 5 and Supplementary Table 5-1). The height of the line shows how large the effect of LUI is relative to all other predictors in the model. Variation in compositional turnover along an environmental or spatial gradient can be seen from the shape of the line, which shows how the effect of a given predictor on compositional turnover varies with the mean level of that predictor. For instance, for LUI, the shape of the line shows how the effect of heterogeneity in LUI varies with mean LUI. We also calculated a bootstrapped P value for each term in the full GDM, using the gdm.varImp function in the gdm library (Supplementary Table 5-2). Additionally we estimated uncertainty for the GDM plots by using 100 bootstraps for each model, each time removing 30% of the plot pairs and then fitting a GDM and extracting the predictions. We then calculated the s.d. of the predictions and added this (\pm) to the fitted line (Extended Data Fig. 6). This is based on the function plotUncertainty in the gdm library.

We fitted GDMs to four measures of β -diversity for each group: turnover (β_{sim}), total β -diversity (Sorensen, $q=0$) and abundance weighted β -diversity ($q=1$ and 2). In each case differences in LUI were fitted as an explanatory variable. To correct for spatial, environmental and other land-use distances, we additionally fitted the spatial distance between plots, differences in pH, differences in nutrients, differences in sLUI and differences in isolation in the model. We also ran GDMs with individual land-use components, that is, with grazing, mowing and fertilization, instead of LUI. These models had the same covariates as the LUI models. For the linear models, we ran the LUI GDMs without soil nutrient levels to test whether some effects of LUI were driven by soil nutrients. In both cases, the effects of LUI were very similar regardless of whether soil nutrients were included or not; this indicates that LUI effects on β -diversity are mostly not caused by LUI homogenizing the soil abiotic environment.

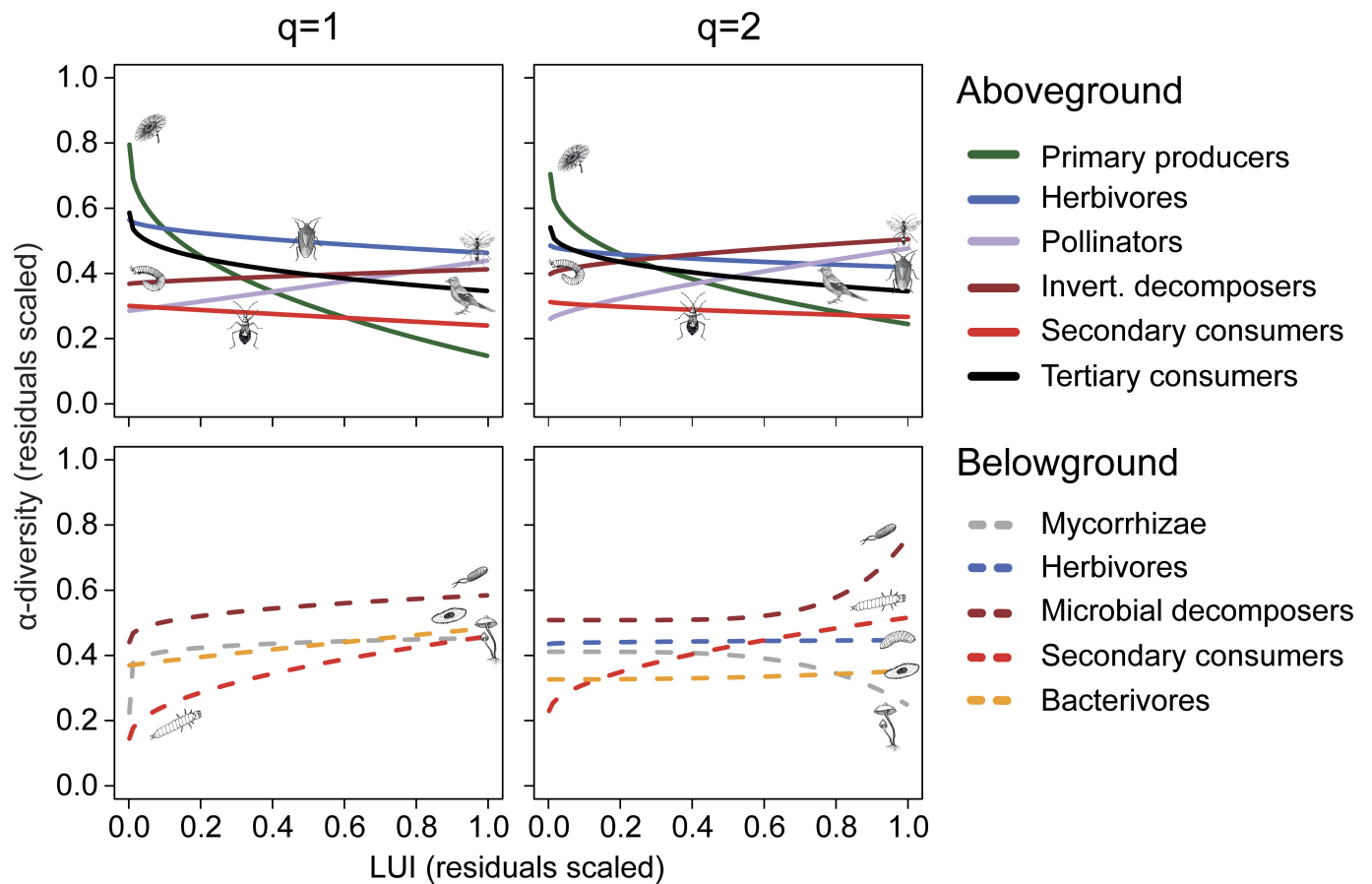
Furthermore, it is possible to estimate the amount of deviance in compositional turnover explained by the GDM. Note that the GDM optimizes the fit between predictors and response variables, so different models can have different response spline shapes. We also determined the proportion of deviance uniquely attributable to land use. We did this by comparing the deviance explained by a GDM containing all of the variables and a GDM with all variables except the difference in LUI between plots. We calculated the unique deviance explained by LUI as the difference in deviance explained between these models. We then converted this to a percentage by dividing by the deviance explained by the full GDM.

Correlation in β -diversity between trophic groups. To test for possible effects of land use on the correlation of β -diversities between trophic groups, we used partial multivariate correlograms and multiple regressions (pmgram and MRM functions in the ecodist library⁶⁹). We correlated β -diversity of different groups (β_{sim} , $q=0$, $q=1$, and $q=2$; see above), and corrected for LUI distances between plots. We corrected for LUI to account for potential shared responses to common environmental drivers. We did this by using the residuals for the matrix correlations between trophic levels. The multiple regressions use permutation tests (999 permutations) of significance for the regression coefficients and for the R^2 values.

To test whether the strength of correlations differed between low and high LUIs, we divided the 105 plots into 52 low (less than median LUI) and 53 high (greater than median LUI) intensity plots and calculated the R^2 -value differences between high and low LUI ($R^2_{\text{high}} - R^2_{\text{low}}$). We then compared these values to a distribution of simulated R^2 -values differences ($n=1,899$) where we randomized the LUI differences between plots. On the basis of this random distribution, we calculated Z scores (standardized effect sizes (SES)) and P values. Significant values thus indicate stronger trophic interactions at lower (or higher) LUI than expected by chance.

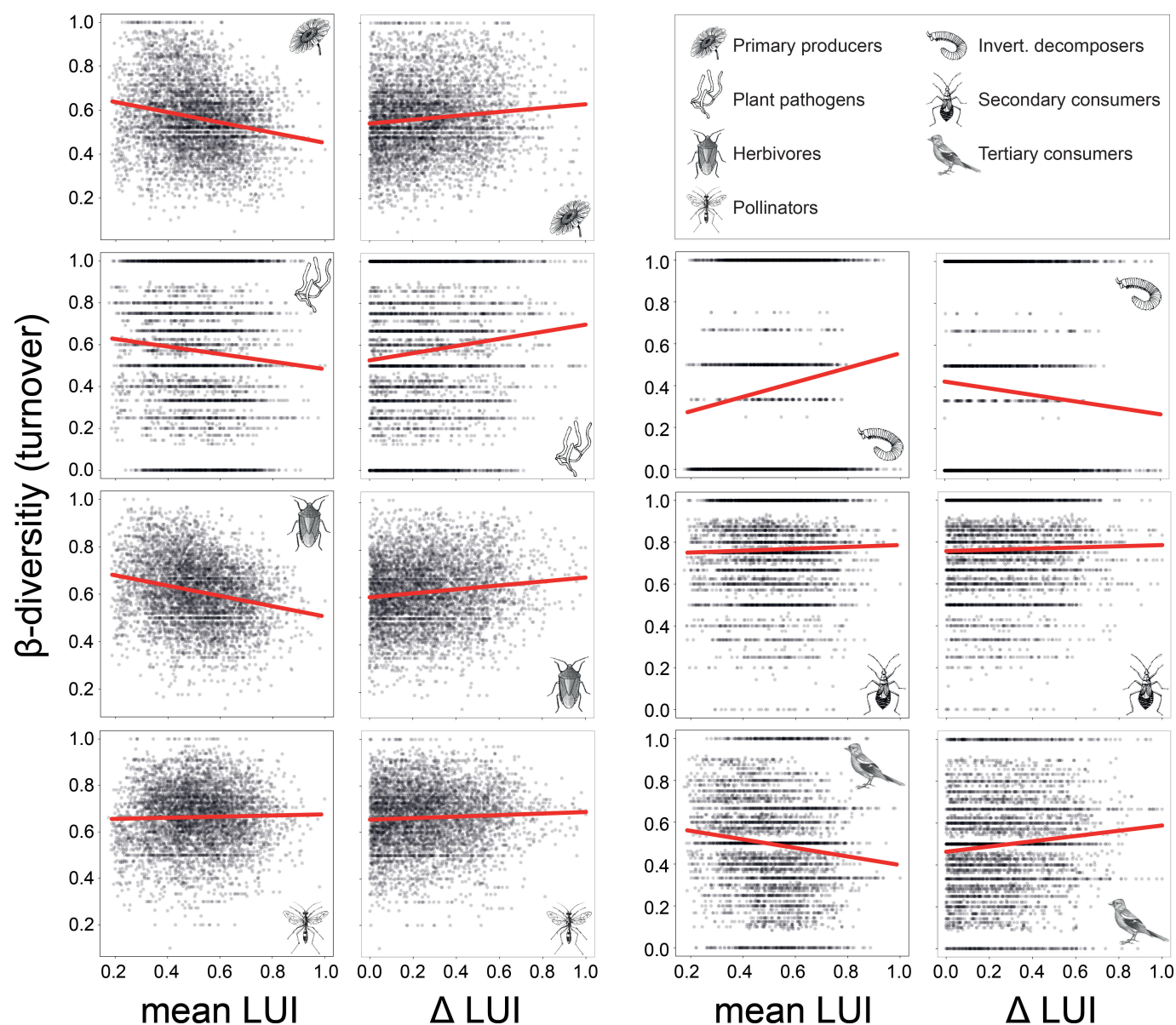
Data availability. The data will become publicly available according to the Rules of Procedure of the German Science Foundation (DFG)-funded Biodiversity Exploratories, that is, five years after completion of the datasets.

29. Fischer, M. *et al.* Implementing large-scale and long-term functional biodiversity research: the Biodiversity Exploratories. *Basic Appl. Ecol.* **11**, 473–485 (2010).
30. Birkhofer, K. *et al.* General relationships between abiotic soil properties and soil biota across spatial scales and different land-use types. *PLoS One* **7**, e43292 (2012).
31. Dray, S. & Dufour, A. B. The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* **22**, 1–20 (2007).
32. *vegan: Community Ecology Package*. R package version 2.3-3. <https://CRAN.R-project.org/package=vegan> (2016).
33. Socher, S. A. *et al.* Interacting effects of fertilization, mowing and grazing on plant species diversity of 1500 grasslands in Germany differ between regions. *Basic Appl. Ecol.* **14**, 126–136 (2013).
34. Socher, S. A. *et al.* Direct and productivity-mediated indirect effects of fertilization, mowing and grazing on grassland species richness. *J. Ecol.* **100**, 1391–1399 (2012).
35. Boch, S., Prati, D., Schöning, I. & Fischer, M. Lichen species richness is highest in non-intensively used grasslands promoting suitable microhabitats and low vascular plant competition. *Biodivers. Conserv.* **25**, 225–238 (2016).
36. Müller, J. *et al.* Impact of land-use intensity and productivity on bryophyte diversity in agricultural grasslands. *PLoS One* **7**, e51520 (2012).
37. Simons, N. K. *et al.* Resource-mediated indirect effects of grassland management on arthropod diversity. *PLoS One* **9**, e107033 (2014).
38. Simons, N. K. *et al.* Effects of land-use intensity on arthropod species abundance distributions in grasslands. *J. Anim. Ecol.* **84**, 143–154 (2015).
39. Weiner, C. N., Werner, M., Linsenmair, K. E. & Blüthgen, N. Land-use impacts on plant–pollinator networks: interaction strength and specialization predict pollinator declines. *Ecology* **95**, 466–474 (2014).
40. Börschig, C. *Effects of land-use intensity in grasslands on diversity, life-history traits and multitrophic interactions* Dr. rer. nat. thesis, Georg-August-Universität (2012).
41. Börschig, C., Klein, A. M., von Wehrden, H. & Krauss, J. Traits of butterfly communities change from specialist to generalist characteristics with increasing land-use intensity. *Basic Appl. Ecol.* **14**, 547–554 (2013). [10.1016/j.baee.2013.09.002](https://doi.org/10.1016/j.baee.2013.09.002)
42. Kempson, D., Lloyd, M. & Ghelardi, R. A new extractor for woodland litter. *Pedobiologia* **3**, 1–21 (1963).
43. Renner, S. C. *et al.* Temporal changes in randomness of bird communities across central Europe. *PLoS One* **9**, e112347 (2014).
44. Rydell, J., Entwistle, A. & Racey, P. A. Timing of foraging flights of three species of bats in relation to insect activity and predation risk. *Oikos* **76**, 243–252 (1996).
45. Denzinger, A., Siemers, B. M., Schaub, A. & Schnitzler, H.-U. Echolocation by the barbastelle bat, *Barbastella barbastellus*. *J. Comp. Physiol. A* **187**, 521–528 (2001).
46. Russo, D. & Jones, G. Identification of twenty-two bat species (Mammalia: Chiroptera) from Italy by analysis of time-expanded recordings of echolocation calls. *J. Zool.* **258**, 91–103 (2002).
47. Obrist, M. K., Boesch, R. & Flückiger, P. F. Variability in echolocation call design of 26 Swiss bat species: consequences, limits and options for automated field identification with a synergetic pattern recognition approach. *Mammalia* **68**, 307–322 (2004).
48. Jung, K., Kaiser, S., Böhm, S., Nieschulze, J. & Kalko, E. K. V. Moving in three dimensions: effects of structural complexity on occurrence and activity of insectivorous bats in managed forest stands. *J. Appl. Ecol.* **49**, 523–531 (2012). [10.1111/j.1365-2664.2012.02116.x](https://doi.org/10.1111/j.1365-2664.2012.02116.x)
49. Fenton, M. B. in *Bat Echolocation Research: Tools, Techniques and Analysis* (eds Brigham, M. *et al.*) 133–140 (Bat Conservation International, 2004).
50. Estrada-Villegas, S., Meyer, C. F. J. & Kalko, E. K. V. Effects of tropical forest fragmentation on aerial insectivorous bats in a land-bridge island system. *Biol. Conserv.* **143**, 597–608 (2010).
51. Lueders, T., Manefield, M. & Friedrich, M. W. Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients. *Environ. Microbiol.* **6**, 73–78 (2004).
52. Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G. & Neufeld, J. D. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl. Environ. Microbiol.* **77**, 3846–3852 (2011).
53. Morris, E. K. *et al.* Land use and host neighbor identity effects on arbuscular mycorrhizal fungal community composition in focal plant rhizosphere. *Biodivers. Conserv.* **22**, 2193–2205 (2013).
54. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
55. Öpiik, M. *et al.* The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New Phytol.* **188**, 223–241 (2010).
56. *R: A language and environment for statistical computing* v. 3.2.2. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/> (2015).
57. Chao, A. *et al.* Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **84**, 45–67 (2014).
58. Chao, A. & Jost, L. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* **93**, 2533–2547 (2012).
59. *iNEXT: iNterpolation and EXTrapolation for species diversity*. R package version 2.0, <http://chao.stat.nthu.edu.tw/blog/software-download> (2014).
60. Jost, L. Entropy and diversity. *Oikos* **113**, 363–375 (2006).
61. Jost, L. Partitioning diversity into independent alpha and beta components. *Ecology* **88**, 2427–2439 (2007).
62. Hill, M. O. Diversity and evenness: unifying notations and its consequences. *Ecology* **54**, 427–432 (1973).
63. Maurer, B. A. & McGill, B. J. in *Biological Diversity: Frontiers in Measurement and Assessment* Vol. 12 (eds Magurran, A. E. & McGill, B. J.) 55–65 (Oxford Univ. Press, 2011).
64. Jost, L., Chao, A. & Chazdon, R. in *Biological Diversity: Frontiers in Measurement and Assessment* Vol. 12 (eds Magurran, A. E. & McGill, B. J.) 66–84 (Oxford Univ. Press, 2011).
65. *vegan: Community Ecology Package*. R package version 2.2-1. (2015).
66. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-125, <http://CRAN.R-project.org/package=nlme> (2016).
67. *lmerPerm: Permutation Tests for Linear Models*. R package version 2.1.0. <https://CRAN.R-project.org/package=lmerPerm> (2016).
68. *gdm: Functions for Generalized Dissimilarity Modeling* v. R-package version 1.1-7 (2016).
69. Goslee, S. C. & Urban, D. L. The ecodist package for dissimilarity-based analysis of ecological data. *J. Stat. Softw.* **22**, 1–19 (2007).

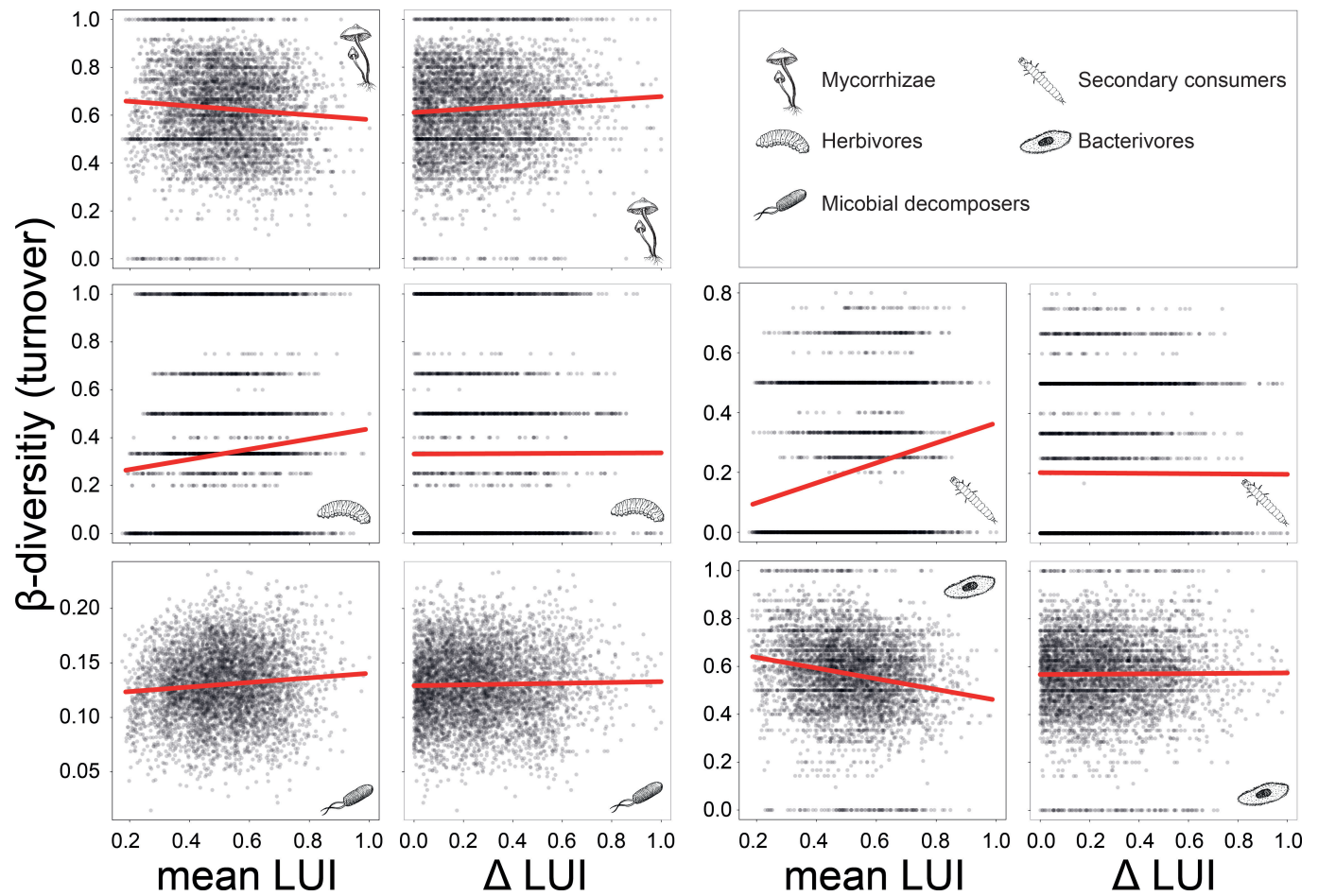


Extended Data Figure 1 | The effect of LUI on higher q -level α -diversity above- and belowground. The partial effect of local LUI comes from a power law model fitted to the exponential Shannon diversity ($q = 1$) and reciprocal Simpson index ($q = 2$) of the seven aboveground (solid lines) and the five belowground trophic groups (dashed lines) ($n = 105$ plots; for more details see Methods). In the model, all parameters of the power law function depended on temporal variation in LUI (sdLUI)

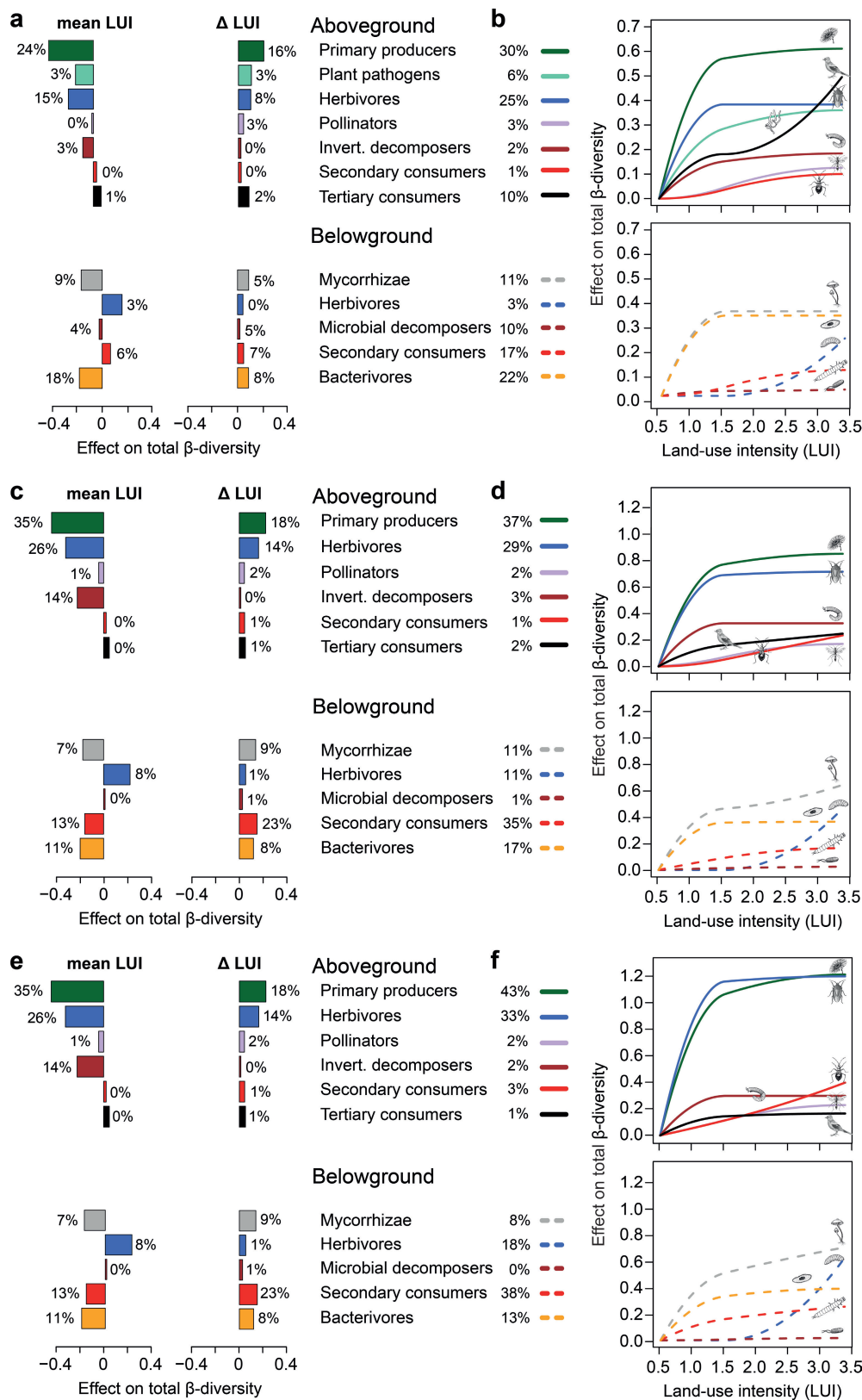
and isolation. LUI effects are plotted at the mean values of these two variables. α -diversity and land-use variables were corrected for differences due to region, pH and soil nutrients, by taking residuals, and were then scaled between 0 and 1. The models for protists ($q = 1$ and $q = 2$) and mycorrhizae ($q = 2$) failed to converge and are therefore not shown. Note that plant pathogens are missing because, for this group, no data on abundance was available.



Extended Data Figure 2 | Effects of LUI on turnover of aboveground species. Scatter plots showing the effects of mean LUI and Δ LUI, between plot pairs ($n = 105$ plots), on the species turnover component of β -diversity for seven aboveground groups. Regression lines show predictions from linear models.

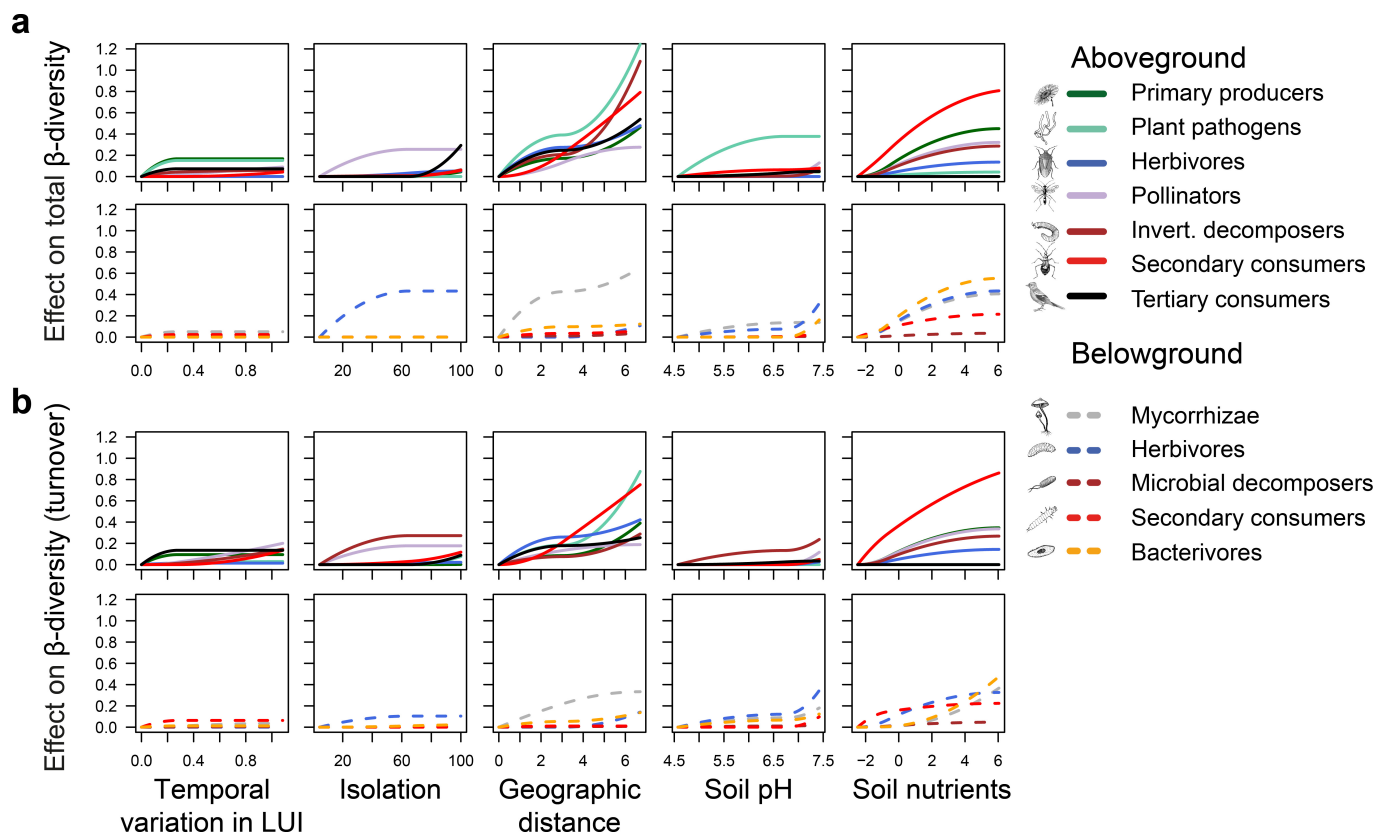


Extended Data Figure 3 | Effects of LUI on turnover of belowground species. Scatter plots showing the effects of mean LUI and Δ LUI, between plot pairs ($n = 105$ plots), on the species turnover component of β -diversity for five belowground groups. Regression lines show predictions from linear models.



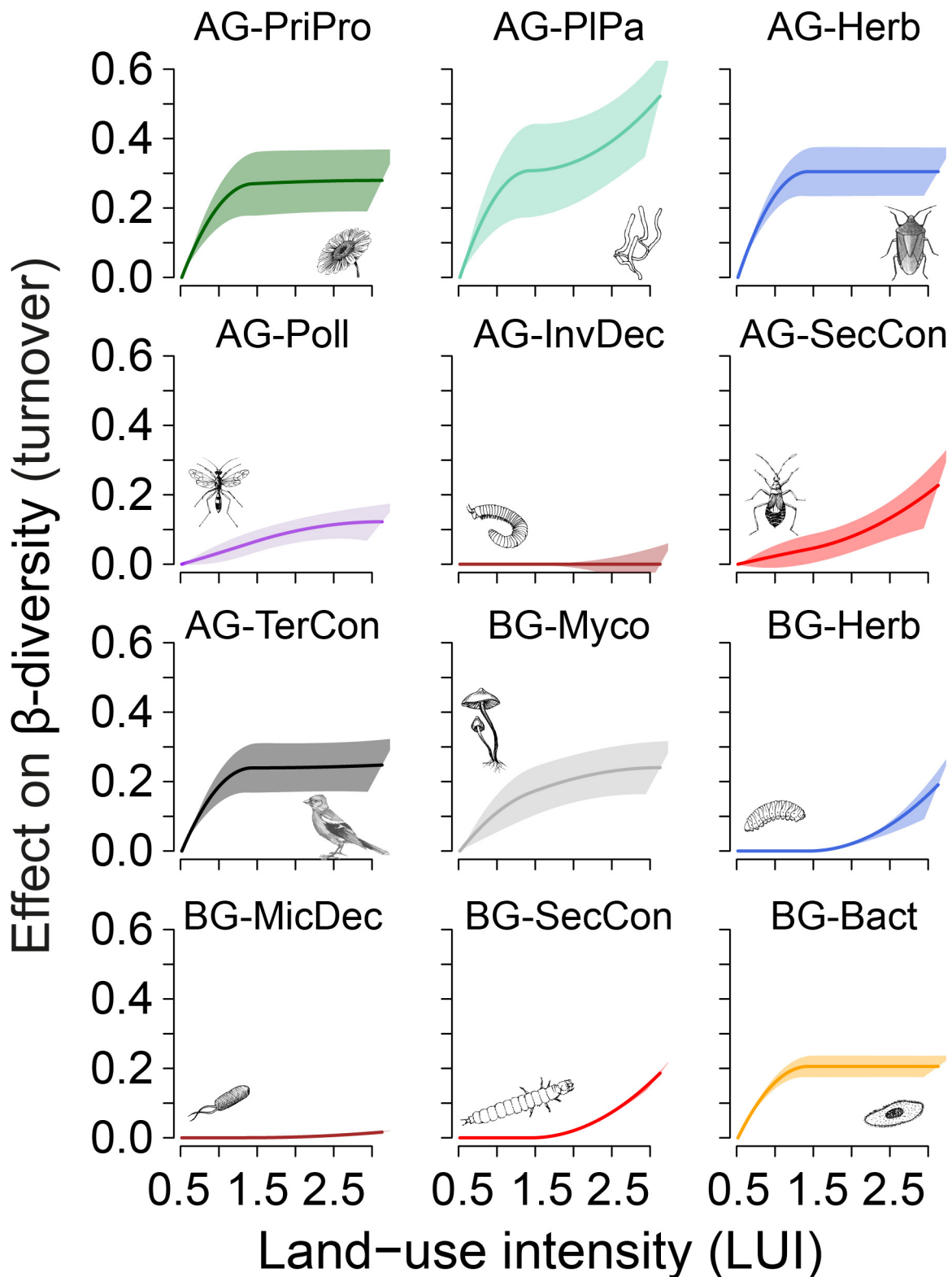
Extended Data Figure 4 | Effects of LUI on total β -diversity above- and belowground. **a, c, e,** Partial effects of mean LUI and Δ LUI, between plot pairs, on total β -diversity (**a**, Sørensen $q = 0$; **c**, Morisita $q = 1$; **e**, Morisita–Horn $q = 2$) for seven aboveground and five belowground groups from linear models. Negative effects of mean LUI indicate that land-use intensification reduces β -diversity. The bars show coefficients from the models. Numbers adjoining bars are the proportion of explained variance uniquely explained by mean LUI or Δ LUI. **b, d, f,** Results from the GDMs are shown for total β -diversity (**b**, Sørensen $q = 0$; **d**, Morisita

$q = 1$; **f**, Morisita–Horn $q = 2$) for the same trophic groups. The figures show the effect of differences in LUI on β -diversity (calculated between all plot pairs). Effects of differences in LUI can vary nonlinearly along the gradient of LUI. Higher maximum curves indicate larger effects of differences in LUI on β -diversity. The values in the legend are the percentage of deviance that is explained uniquely by LUI. Effects of both linear models and GDMs are corrected for other drivers of β -diversity, and response and explanatory variables are scaled to allow comparisons across trophic groups ($n = 105$ plots; for details see Methods).



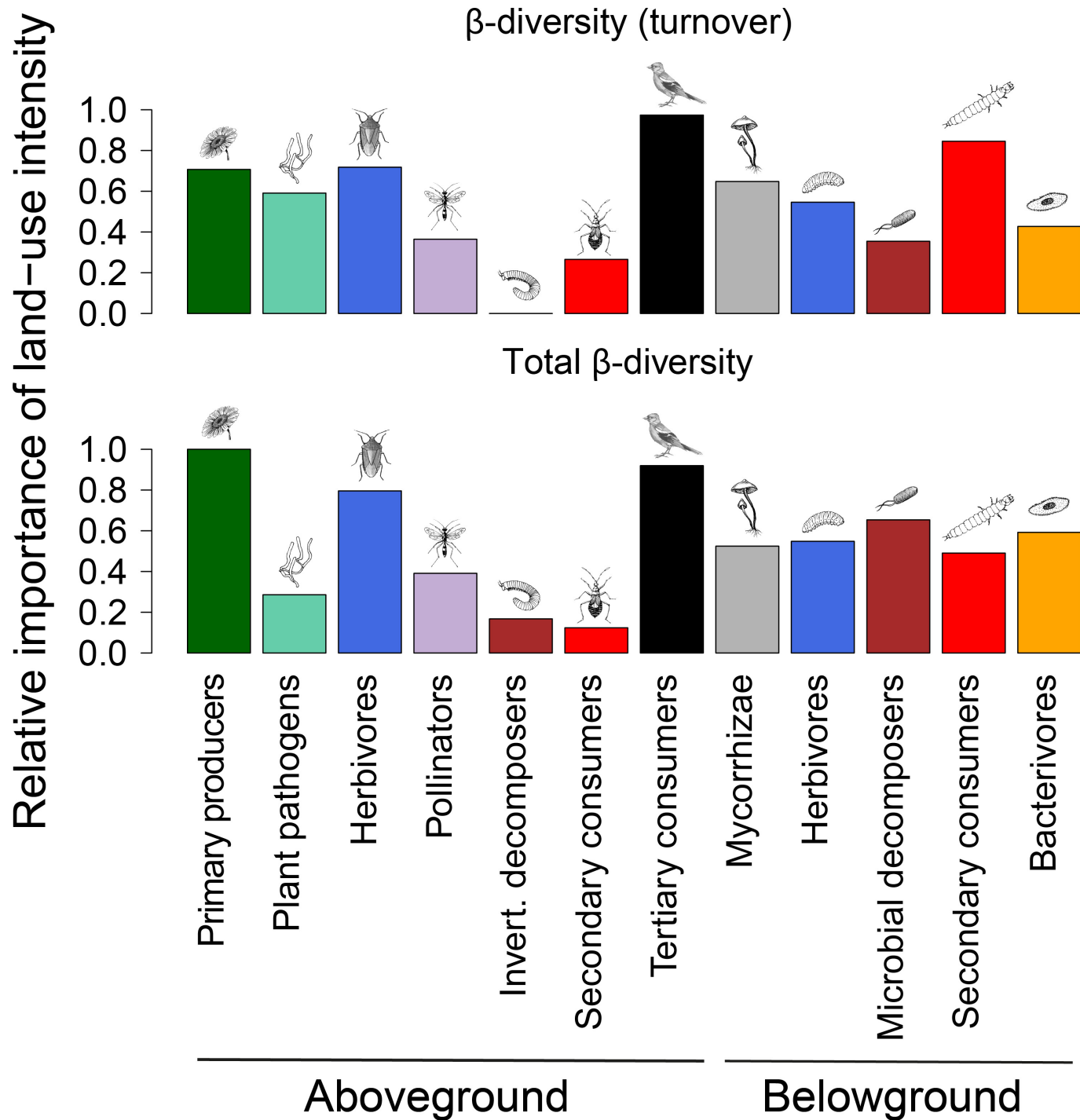
Extended Data Figure 5 | Partial effects of geographic and environmental distances and temporal variation in LUI on β -diversity above- and belowground. **a, b,** Results from GDMs are shown for seven aboveground and five belowground groups, with total β -diversity measured as the Sørensen index β_{sor} (**a**) or as the species turnover component β_{sim} (**b**). The figures show the effect of differences in each of the five variables on β -diversity (calculated between all plot pairs;

$n = 105$ plots). Effects of differences in each explanatory variable can vary nonlinearly along the gradient of that variable and each is corrected for all other variables in the model. Higher maximum curves indicate larger effects of differences in a given variable on β -diversity. Soil nutrients refer to the scores of the first PCA axis. Temporal variation in LUI is shown as s.d. Geographic distance has to be multiplied by 100 km.

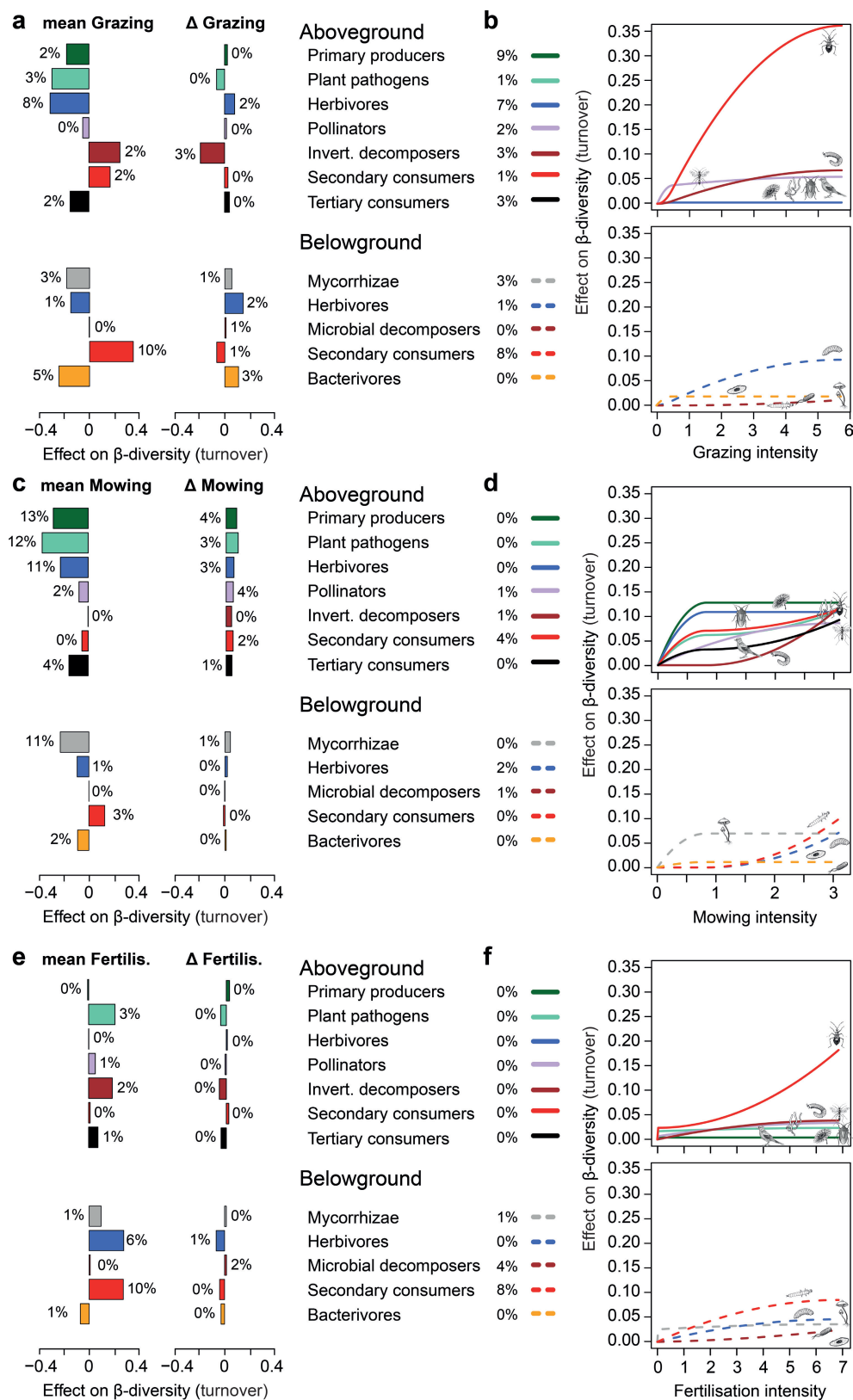


Extended Data Figure 6 | Uncertainty of effects of LUI on β -diversity above (AG) and belowground (BG). The uncertainty is calculated on the basis of 100 bootstraps for each model, each time removing 30% of the plot pairs, then fitting a GDM and extracting the predictions. Predictions are shown as fitted lines and s.d. Uncertainty is shown for all seven above- and

five belowground trophic groups based on species turnover β_{sim} ($n = 105$ plots). PriPro, primary producers; PIPa, plant pathogens; Herb, herbivores; Poll, pollinators; InvDec, invertebrate decomposers; SecCon, secondary consumers; TerCon, tertiary consumers; Myco, Mycorrhizae; MicDec, microbial decomposers; Bact, bacterivores.

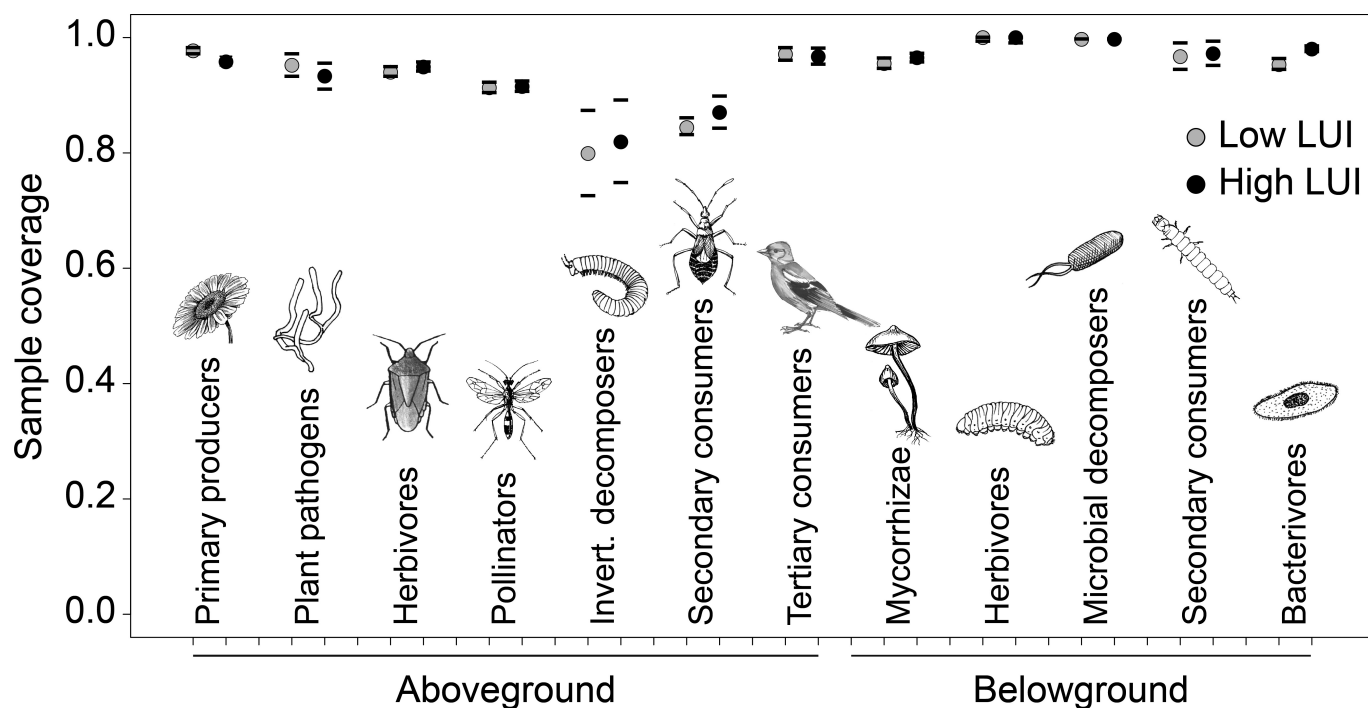


Extended Data Figure 7 | The relative importance of LUI as a driver of β -diversity. The bar plot shows the importance of LUI (in terms of total effect size) relative to the most important variable in the GDM. Results are shown for each trophic group, for the species turnover component (β_{sim}) and total β -diversity (Sørensen index) ($n = 105$ plots).



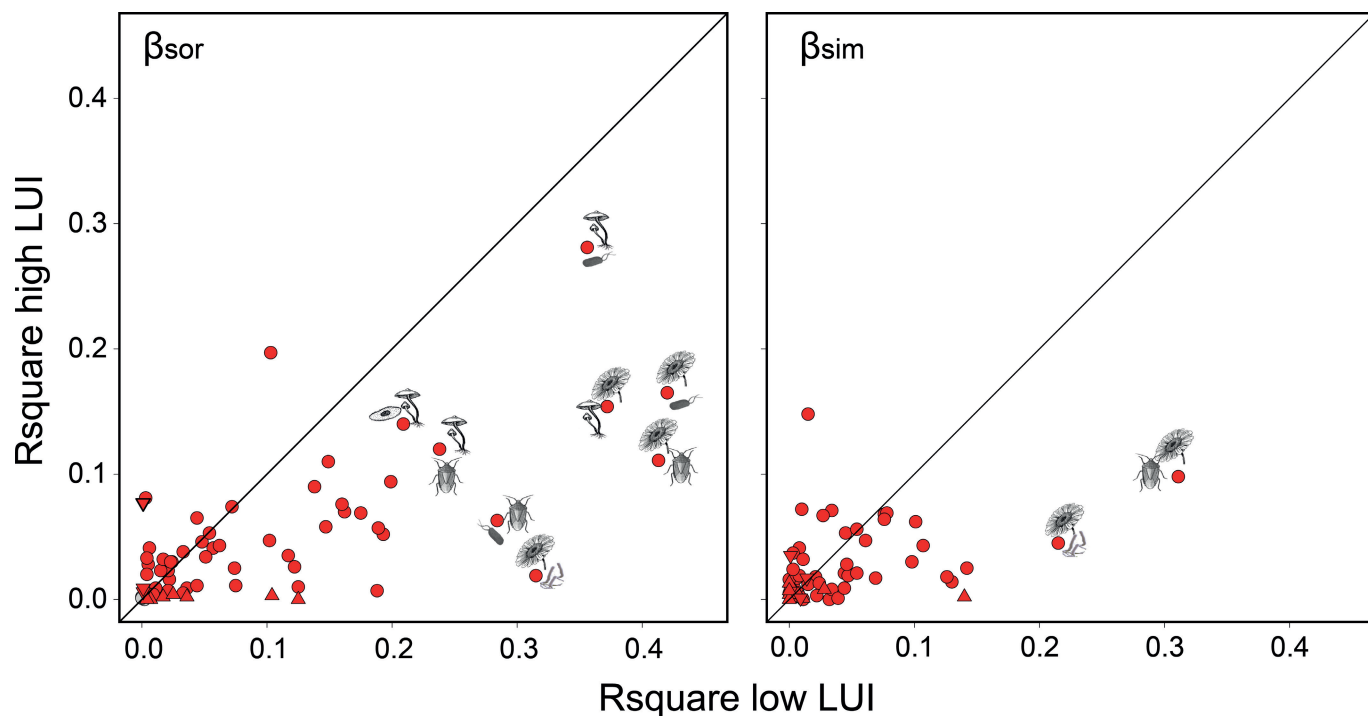
Extended Data Figure 8 | Effects of single land-use components on β -diversity above- and belowground. **a, c, e,** Partial effects of minimum LUI (min LUI) and Δ LUI between plot pairs ($n = 105$ plots), on the species turnover component of β -diversity (β_{sim}) for seven aboveground and five belowground groups based on linear models. Negative effects of minimum LUI indicate that land-use intensification reduces β -diversity. The bars show coefficients from the models. Numbers adjoining bars are the proportion of the total explained variance that is uniquely explained by minimum LUI or Δ LUI among plot pairs, on the basis of hierarchical

partitioning. **b, d, f,** Results from GDMs are shown for the turnover component β_{sim} for the same trophic groups. The figures show the effect of Δ LUI on β -diversity (calculated between all plot pairs). Effects of Δ LUI can vary nonlinearly along the gradient of LUI. Higher maximum curves indicate larger effects of Δ LUI on β -diversity. The values in the legend are the percentage of deviance that is explained uniquely by LUI. Effects of both linear models and GDMs are corrected for other drivers of β -diversity, and response and explanatory variables are scaled to allow comparisons across trophic levels (see Methods).



Extended Data Figure 9 | Sample coverage of above- and belowground trophic groups based on species incidences. Sample coverage was calculated for low (52 plots) and high (53) LUI plots based on refs 57, 58. Coverage is defined as the proportion of the total number of individuals

in an assemblage that belong to species represented in the sample, and is therefore a measure of sampling completeness. Means and 95% confidence intervals based on 200 bootstraps are shown.



Extended Data Figure 10 | The effect of LUI on the correlation between the β -diversities of different trophic groups. Each dot represents the correlation (R^2) between two trophic groups. Correlations are R^2 values from matrix regressions between β -diversity values of different groups (corrected for effects of differences in LUI on β -diversity). Significant correlations ($P < 0.05$) are marked in red. Upward and downward triangles

indicate significance under low or high LUI only. Interactions with R^2 values higher than 0.2 in one of the LUI-categories are illustrated by icons. β -diversity was calculated as the Sørensen index (β_{sor} , total β -diversity) and as the species turnover component (β_{sim}) ($n = 105$ plots). For statistical details see Supplementary Information Section 5.

Mitochondrial replacement in human oocytes carrying pathogenic mitochondrial DNA mutations

Eunju Kang^{1,2†}, Jun Wu³, Nuria Marti Gutierrez^{1,2}, Amy Koski^{1,2}, Rebecca Tippner-Hedges^{1,2}, Karen Agaronyan⁴, Aida Platero-Luengo³, Paloma Martinez-Redondo³, Hong Ma^{1,2}, Yeonmi Lee^{1,2†}, Tomonari Hayama^{1,2}, Crystal Van Dyken^{1,2}, Xinjian Wang⁵, Shiyu Luo⁵, Riffat Ahmed^{1,2}, Ying Li^{1,2}, Dongmei Ji^{1,6}, Refik Kayali⁷, Cengiz Cinnioglu⁷, Susan Olson⁸, Jeffrey Jensen⁹, David Battaglia⁹, David Lee⁹, Diana Wu⁹, Taosheng Huang⁵, Don P. Wolf^{1,2}, Dmitry Temiakov⁴, Juan Carlos Izpisua Belmonte³, Paula Amato⁹ & Shoukhrat Mitalipov^{1,2,9,10,11}

Maternally inherited mitochondrial (mt)DNA mutations can cause fatal or severely debilitating syndromes in children^{1–3}, with disease severity dependent on the specific gene mutation and the ratio of mutant to wild-type mtDNA (heteroplasmy) in each cell and tissue⁴. Pathogenic mtDNA mutations are relatively common, with an estimated 778 affected children born each year in the United States⁵. Mitochondrial replacement therapies or techniques (MRT) circumventing mother-to-child mtDNA disease transmission involve replacement of oocyte maternal mtDNA^{6–8}. Here we report MRT outcomes in several families with common mtDNA syndromes. The mother's oocytes were of normal quality and mutation levels correlated with those in existing children. Efficient replacement of oocyte mutant mtDNA was performed by spindle transfer⁸, resulting in embryos containing >99% donor mtDNA. Donor mtDNA was stably maintained in embryonic stem cells (ES cells) derived from most embryos. However, some ES cell lines demonstrated gradual loss of donor mtDNA and reversal to the maternal haplotype. In evaluating donor-to-maternal mtDNA interactions, it seems that compatibility relates to mtDNA replication efficiency rather than to mismatch or oxidative phosphorylation dysfunction. We identify a polymorphism within the conserved sequence box II region of the D-loop as a plausible cause of preferential replication of specific mtDNA haplotypes. In addition, some haplotypes confer proliferative and growth advantages to cells. Hence, we propose a matching paradigm for selecting compatible donor mtDNA for MRT.

Because mitochondrial disease can be attributed to gene mutations in mtDNA and/or nuclear DNA⁹, an important clinical challenge is to first confirm pathogenic mtDNA mutations in families eligible for MRT. We recruited four families diagnosed with Leigh syndrome and one with MELAS (mitochondrial encephalomyopathy with lactic acidosis and stroke-like episodes) and conducted genetic testing to confirm maternal inheritance (Extended Data Fig. 1a). We collected DNA from blood, skin fibroblasts and/or urine from children and mothers, and performed whole mtDNA sequencing. The first Leigh syndrome family had an affected 2-year-old child with a homoplasmic T8993G substitution encoding the ATPase 6 gene (*MT-ATP*) of mtDNA in both blood and skin fibroblast samples while her 22-year-old mother had the same mutation with 70% heteroplasmy in blood and 100% in skin fibroblasts (Fig. 1a, left). In the second related Leigh syndrome family, an affected 2.5-year-old child carried the same T8993G mutation at 95%

heteroplasmy in blood and 100% in skin fibroblasts, while in a second 1-year-old asymptomatic sibling, the mutation load was 50% in blood and 62% in skin fibroblasts. Their 23-year-old mother, who is the elder sister of the subject from the first family, carried this mutation at 13% in blood and 16% in skin fibroblasts (Fig. 1a, left). In the third Leigh syndrome family, the affected 12-year-old boy had a G13513A substitution affecting the *MT-ND5* gene at 56%, 86% and 97% heteroplasmy in blood, skin fibroblasts and urine, respectively. His asymptomatic 19-year-old brother carried the same mutation at 10%, 14% and 23% heteroplasmy levels in blood, skin fibroblasts and urine, respectively. Their 36-year-old mother also harboured the mutation at 3%, 98% and 39% levels in blood, skin fibroblasts and urine, respectively (Fig. 1a, middle). A fourth family presented with a 1-year-old child diagnosed with Leigh syndrome. However, genetic screening did not reveal any pathogenic mtDNA mutations in the child or mother. The last family was from a large, well-studied MELAS pedigree carrying a pathogenic A3243G mutation in the *MT-TL1* encoding tRNA leucine¹⁰ (Extended Data Fig. 1b). The 32-year-old mother carried the mutation at 14%, 47% and 35% in blood, skin fibroblasts and urine, respectively. All three of her children inherited this mutation (Fig. 1a, right). Thus, pathogenic mtDNA mutations were implicated in four of the five families, highlighting the importance of genetic testing for maternally inherited mtDNA diseases before MRT. Moreover, because heteroplasmy levels may vary among different tissues, it is critical to sample and test blood, skin and urine, in both mothers and children¹¹.

We also screened 11 healthy volunteer oocyte donors and confirmed that they did not carry inherited pathogenic mtDNA mutations. The mtDNA sequences and corresponding haplotypes were used for matched MRT combinations (Supplementary Table 1).

Women with mtDNA disease display live birth rates comparable to the general population⁵, suggesting normal fertility. We studied ovarian response to gonadotropin stimulation and oocyte recovery in women carrying pathogenic mtDNA mutations (carriers). Anti-Müllerian hormone levels, a measure of ovarian reserve, were lower in carriers than in healthy donors (1.1 compared to 4.8 ng ml⁻¹). Antral follicle count (AFC) was also lower in carriers compared to the healthy donors (10.3 compared to 22.3). In addition, the duration of controlled ovarian stimulation was about one day longer in the carriers and their peak blood oestradiol (E2) level before human chorionic gonadotropin (hCG) administration also trended lower. Finally, the total number of oocytes and the number of mature metaphase II oocytes retrieved were

¹Center for Embryonic Cell and Gene Therapy, Oregon Health & Science University, 3303 SW Bond Avenue, Portland, Oregon 97239, USA. ²Division of Reproductive & Developmental Sciences, Oregon National Primate Research Center, Oregon Health & Science University, 505 NW 185th Avenue, Beaverton, Oregon 97006, USA. ³Gene Expression Laboratory, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. ⁴Department of Cell Biology School of Osteopathic Medicine, Rowan University, 2 Medical Center Drive, Stratford, New Jersey 08084, USA. ⁵Division of Human Genetics, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, Ohio 45229, USA. ⁶Reproductive Medical Centre, Anhui Medical University, No 218, Jixi Rd, Shushan District, Hefei, Anhui 230022, China. ⁷IviGen Los Angeles, 406 Amapola Avenue, Suite 215, Torrance, California 90501, USA. ⁸Research Cytogenetics Laboratory, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239, USA. ⁹Division of Reproductive Endocrinology, Department of Obstetrics and Gynecology, Oregon Health & Science University, 3303 SW Bond Avenue, Portland, Oregon 97239, USA. ¹⁰Knight Cardiovascular Institute, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239, USA. ¹¹Department of Biomedical Engineering, Oregon Health & Science University, 3303 SW Bond Avenue, Portland, Oregon 97239, USA. [†]Present address: Stem Cell Center, ASAN Institute for Life Sciences, ASAN Medical Center, Olympic-ro 43-gil, Songpa-gu, Seoul 138-736, South Korea.

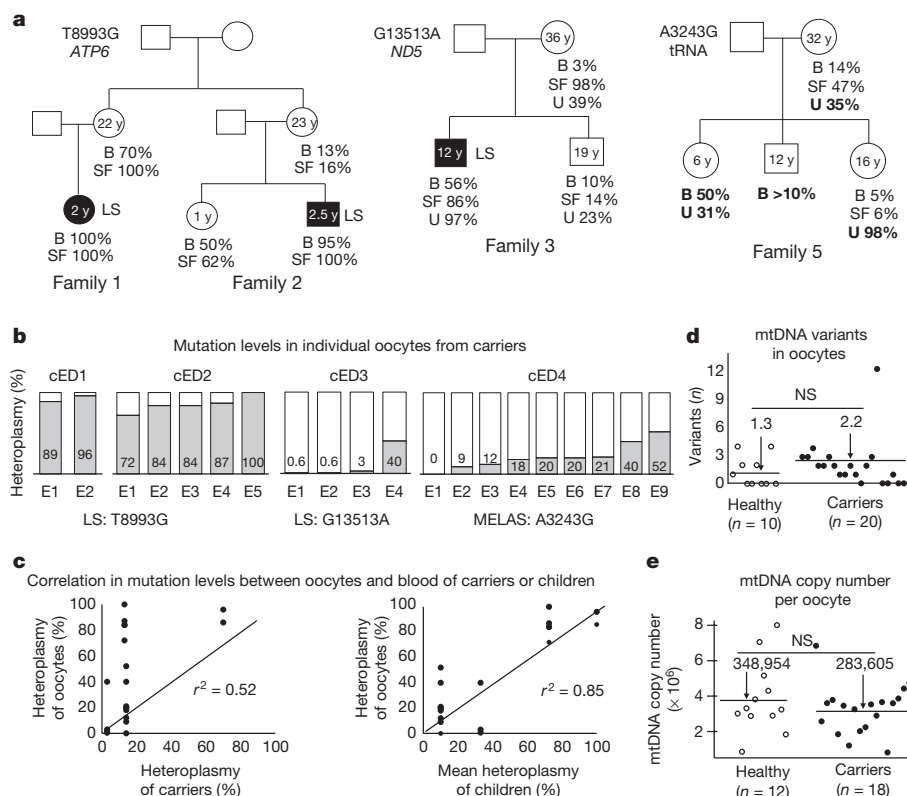


Figure 1 | Characteristics of families and retrieved oocytes in women carrying pathogenic mtDNA mutations (carriers) or wild-type mtDNA (healthy). **a**, Somatic tissue heteroplasmy levels in families (see also Extended Data Fig. 1b). Heteroplasmy levels obtained from previous reports¹⁰ are shown in bold. In all families, the heteroplasmy levels for mtDNA mutations were diverse among tissues and individuals, and associated with clinical disease. See also Supplementary Table 2. Black filled, diagnosed with a mitochondrial disease. LS, Leigh syndrome. y, years-old. Square, male; circle, female. B, blood; SF, skin fibroblasts; U, urine. **b**, Individual oocytes harboured different heteroplasmy levels. cED, carrier egg donor. **c**, The mean heteroplasmy in oocytes correlated

with levels in existing children rather than with maternal levels. Heteroplasmy levels of carriers and children were measured in blood. r^2 , Pearson's r values. **d**, Comparable mtDNA variants between oocytes with healthy mtDNA and those carrying mutations ($P > 0.05$). **e**, Mean mtDNA copy number per oocyte in healthy controls was comparable to those carrying pathogenic mtDNA ($P > 0.05$). Karyoplasts were isolated from carrier oocytes for spindle transfer, except oocytes recovered from cED2 who exhibited premature luteinization. Oocytes from healthy egg donors were mature metaphase II. n, the number of oocytes; the number in each group is a mean; NS, not significant (**d** and **e**; analysed by t -test). Source data files are available for **c–e**.

also lower in carriers (5.8 versus 16.6 and 3.8 versus 13.2, respectively) (Extended Data Fig. 2a–g). Of note, one carrier (cED 2) exhibited premature luteinization, as evidenced by increased progesterone levels before hCG administration (Extended Data Fig. 2h). Therefore, we only performed mutation analyses on retrieved atretic oocytes. Although the numbers in our cohort were low¹², the results suggest that mtDNA mutations could lead to a reduced ovarian response. Older age and a higher body mass index of carriers could also have affected the outcome^{13,14}. Another potential contributing factor may be the long-term hormonal contraception these women were on before ovarian stimulation.

After metaphase II spindle (nuclear genome) removal, cytoplasts (mtDNA) were used for sequencing. In the first carrier, two recovered oocytes displayed a T8993G mutation at 89% and 96% levels. In the second sibling carrier, heteroplasmy levels in five atretic oocytes ranged from 72% to 100%. In four oocytes from the third Leigh syndrome carrier, G13513A mutation levels in three eggs were low (0.6–3%) while the fourth was 40%. Finally, in nine oocytes from the A3243G MELAS carrier, no mutation was detected in one egg while the others carried 9–52% (Fig. 1b). Mean heteroplasmy levels in the blood of sibling children correlated highly with those in oocytes^{4,15} ($r^2 = 0.85$ compared to $r^2 = 0.52$ between mutation levels in the carrier's blood and oocytes, Fig. 1c). Next, we analysed whole mtDNA sequences in carrier oocytes to screen for secondary mtDNA mutations and compared to oocytes from healthy donors. Carrier oocytes contained secondary heteroplasmic mtDNA variants, however the average number per oocyte was not

significantly different from healthy controls (2.2 versus 1.3; $P > 0.05$) (Fig. 1d). Some of these variants were present in several oocytes from the same carrier and were also found in the carrier's blood, skin or in sibling children, suggesting recurring germline mutations. However, these variants were non-coding D-loop mutations (Supplementary Table 2). Others were found only in one oocyte within the group indicative of *de novo* mutations (45%). The majority (83%) of these secondary variants were low-level heteroplasmic (<15%) mutations (Extended Data Fig. 2i). We also measured mtDNA copy number as an indirect measure of oocyte quality and found no significant differences between healthy and carrier oocytes (mean = 348,954 compared to 283,605, respectively) (Fig. 1e). A low mtDNA copy number may play a role in diminished ovarian reserve¹⁶ but we found no correlation between oocyte copy number and anti-Müllerian hormone levels (Extended Data Fig. 2j).

Meiotic spindles (karyoplasts containing carryover maternal mtDNA) recovered from carrier oocytes ($n = 13$) were transferred into enucleated donor oocytes (cytoplasts; donor mtDNA) (Fig. 2a) while controls involved spindle transfer (ST) between healthy oocytes ($n = 36$) with preselected mtDNA haplotypes (Extended Data Fig. 3a). Spindle transfer oocytes, along with intact (non-manipulated) controls, were then fertilized by intracytoplasmic sperm injection and cultured to blastocysts. High fertilization rates, comparable to controls, were observed in both spindle transfer groups (Fig. 2a and Extended Data Fig. 3b, c). Subsequent development of diploid carrier spindle transfer zygotes to the blastocyst stage was 75%, similar to controls. Similar

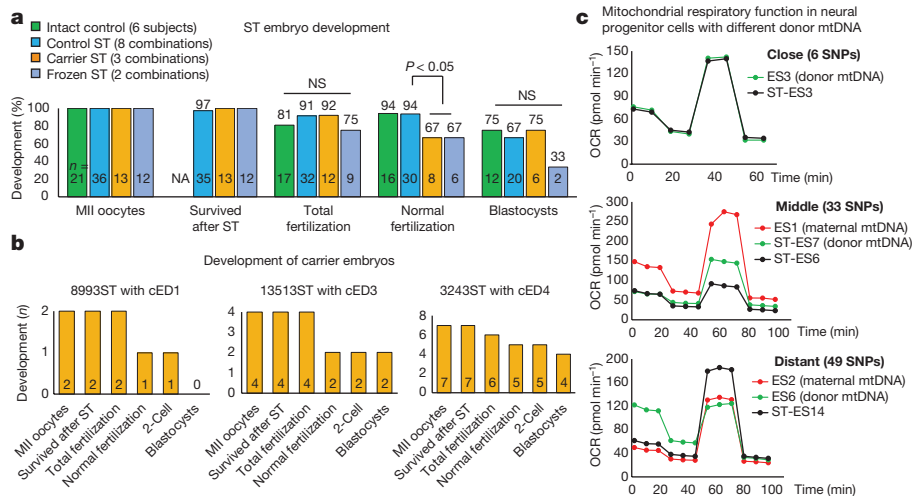


Figure 2 | Spindle transfer outcomes, embryo development and mitochondrial function. **a**, Spindle transfer (ST) embryo development following various experimental manipulations. Similar developmental outcomes were observed among groups, except normal fertilization, which was significantly lower in carrier and frozen spindle transfer groups ($P < 0.05$; one-way ANOVA). Numbers inside bars, number of oocytes or embryos; numbers on the top of bars, percentages. Combinations, mixtures of different mtDNA haplotypes between karyoplast and cytoplasm; intact control, unmanipulated embryos, fertilized by intracytoplasmic sperm injection (ICSI). NA, not applicable. **b**, Carrier spindle transfer embryos

developmental patterns were observed among carriers with different mutation types (8993ST, 13513ST and 3243ST) (Fig. 2b). To address controlled ovarian stimulation asynchrony, we conducted spindle transfer between fresh and vitrified oocytes (frozen spindle transfer). The overall fertilization rate using vitrified oocytes was comparable to fresh spindle transfer, while formation of normal diploid zygotes was reduced (Fig. 2a). No differences in fertilization were noted between combinations involving vitrified karyoplasts versus vitrified cytoplasm (Extended Data Fig. 3d).

Since miscommunication between nuclear and mitochondrial genomes may induce secondary metabolic dysfunction¹⁷, we analysed spindle transfer embryo development as a function of donor mtDNA sequence distance, as measured indirectly by the number of single nucleotide polymorphism (SNP) differences. Spindle transfer oocytes and embryos were grouped based on the SNP differences between the original (maternal) and donor oocyte mtDNA, ranging from close (6 SNPs) to middle (33 SNPs) to distant (57 SNPs) (Extended Data Fig. 3a and Supplementary Table 1). Fertilization and blastulation rates were similar among all three spindle transfer groups, suggesting that embryo development was not compromised (Extended Data Fig. 3e). Next, we established ES cell lines from spindle transfer blastocysts and differentiated them into neural progenitor cells (NPCs), cardiomyocytes and teratomas. Comparable differentiation efficiencies, mitochondrial respiratory chain enzyme activity and oxygen consumption rates were found between groups (Fig. 2c and Extended Data Fig. 4, 5). Thus, MRT in carrier oocytes supported fertilization and blastocyst development similarly to controls. Also, ES cell derivation and differentiation were unaffected by donor mtDNA genetic backgrounds. However, it is still possible that miscommunication between certain combinations of nuclear and mitochondrial genomes may occur, independent of SNP differences.

We biopsied and analysed expanded blastocysts with a comparative genome hybridization array. Aneuploidy rates were similar between spindle transfer and intact controls (Extended Data Fig. 6a). G-banding karyotype analysis in spindle transfer and control ES cells also did not reveal any significant differences (Extended Data Fig. 6b). Finally, we examined copy number variations (CNVs) in selected spindle transfer ES cell lines for subchromosomal abnormalities. *De novo* CNVs were

with different mtDNA mutations showed similar development patterns. Numbers inside bars, number of oocytes or embryos. MII, mature metaphase II. **c**, Neural progenitor cells (NPCs) derived from spindle transfer ES cells with 6 and 49 SNP differences in donor mtDNA had similar oxygen consumption rates (OCR) to non-manipulated controls. However, NPCs with 33 SNPs showed significantly reduced oxygen consumption rates compared to controls with either donor or maternal mtDNA ($P < 0.05$, one-way ANOVA). See Extended Data Fig. 4a. Seahorse assay results were based on 16 technical replicates. SNPs, single nucleotide polymorphisms. Source data files are available for **a**, **b**.

detected in both spindle transfer and intact controls but were deemed of uncertain clinical significance¹⁸ (Extended Data Table 1). In summary, a total of six spindle transfer blastocysts were produced from four ovarian stimulation cycles in carriers (Extended Data Table 2). These outcomes were comparable to controls.

A small amount of maternal mtDNA carry-over is common during spindle transfer, resulting in low heteroplasmy in human embryos and in non-human primate offspring (typically below 2%)^{19,20}. Here, all examined spindle transfer zygotes and cleaving embryos ($n = 22$) contained more than 99% donor mtDNA (Table 1). Similar outcomes were observed in 13 out of 15 ES cell lines (87%) established from control spindle transfer blastocysts, regardless of donor mtDNA (Supplementary Table 3). However, two sibling ES cell lines (ST-ES7 and ST-ES8), generated by a combination of maternal U5a and donor H1b mtDNA (33 SNPs), displayed high levels of maternal mtDNA (81% and 94%, respectively) (Table 1). Extended passaging resulted in a complete loss of donor mtDNA and a return to the original maternal mtDNA haplotype. By contrast, another sibling ES cell line generated from maternal U5a, but with donor V3 (ST-ES9; 33 SNPs) maintained donor mtDNA.

Among the carrier spindle transfer ES cell lines, 3243ST-ES1 also demonstrated 20% maternal H49 mtDNA (Table 1). However, this maternal haplotype did not include the A3243G mutation (Supplementary Table 3). Maternal mtDNA levels in this cell line gradually increased during extended culture to 90% at passage 8 and homoplasmy at passage 10 (Fig. 3c). However, another sibling 3243ST-ES2 generated by the same maternal and donor mtDNA combinations did not show a reversal (Table 1 and Supplementary Table 3).

We expanded our study to include eight ES cell lines derived by somatic cell nuclear transfer (SCNT, NT-ES cells) that also carry donor oocyte mtDNA^{21–23} (Table 1 and Supplementary Table 4). We identified that NT-ES8 displayed a gradual increase in maternal (somatic) mtDNA from 19% at passage 2 to 100% at passage 10 (Table 1 and Extended Data Fig. 7a). We also investigated whether the reversal is specific to undifferentiated ES cells or could occur during differentiation. Several reversed and non-reversed ES cells were differentiated *in vitro* into NPCs and cardiomyocytes, and *in vivo* into teratomas (Extended Data Fig. 5). Maternal mtDNA levels were measured before and after differentiation.

Table 1 | Donor mtDNA in pre-implantation MRT embryos and ES cell lines

Treatment	Carryover maternal mtDNA %					Maternal haplotype	Donor haplotype	SNP differences		
	Pre-implantation embryos		ES cells					Total	Coding*	D-loop
	Embryo stage	%	Lines	Passage 2–3	Passage 9–10					
Control ST	NT		ST-ES1†	≤1	≤1	H56	H2a	6	4	1
			ST-ES2†	≤1	≤1	H2a	H56	6	4	1
			ST-ES3	≤1	≤1	H2a	H56	6	4	1
			ST-ES4†	≤1	≤1	H44a	H13a	25	7	10
			ST-ES5	≤1	≤1	H1b	U5a	33	11	9
			ST-ES6	≤1	≤1	H1b	U5a	33	11	9
	4-cell	0.7	ST-ES7	81±9	100	U5a	H1b	33	11	9
	8-cell	0.9								
	10-cell	0.4								
	10-cell	0.4	ST-ES8	94±4	100	U5a	H1b	33	11	9
	1-cell	0.4	ST-ES9	≤1	≤1	U5a	V3	33	11	9
	2-cell	0.5								
	8-cell	0.5								
	Morula	0.4	ST-ES10	≤1	≤1	V3	U5a	33	11	9
	1-cell	0.5†	ST-ES11†	≤1	≤1	H1e	D1f	44	15	14
	NA	0.9†	ST-ES12†	≤1	≤1	H1e	D1f	44	15	14
		1.6†	ST-ES13†	≤1	≤1	H1e	D1f	44	15	14
		0.0†								
	1-cell	0.5	ST-ES14	≤1	≤1	D4a	A2g	49	18	16
	Morula	0.2								
	Morula	0.2								
	1-cell	0.2								
	1-cell	0.5								
	1-cell	0.5	ST-ES15	≤1	≤1	A2g	D4a	49	18	16
	1-cell	0.5								
	Carrier ST	1-cell	0.8	NA			H1b	H6a	21	6
3-cell		0.5								
3-cell		0.6								
NT			13513ST-ES	≤1	≤1	T2b	T2	22	3	5
		1-cell	0.5	3243ST-ES1	20±13	100	H49	B2k	32	7
8-cell		0.8	3243ST-ES2	≤1	≤1	H49	B2k	32	7	14
12-cell		0.5								
Morula	0.5									
SCNT	NT		NT-ES1		≤1	H1b	H56	13	5	7
			NT-ES2		≤1	H1b	H56	13	5	7
			NT-ES3		≤1	H1b	H56	13	5	7
			NT-ES4		≤1	H1b	H56	13	5	7
			NT-ES5		≤1	F1a	D1a	47	16	11
			NT-ES6		≤1	F1a	H1b	38	13	13
			NT-ES7	≤1	≤1	X2c	U5a	40	9	16
			NT-ES8	19±9	100	X2c	D4a	39	12	13

*SNPs resulting in amino acid change in protein-coding genes and residing in RNA genes. NT, not tested. NA, not applicable. ST, spindle transfer; SCNT, somatic cell nuclear transfer.
†Ref. 19.

We observed that maternal mtDNA levels in the reversed 3243ST-ES1 reduced to undetectable levels in some tissues but were slightly elevated in other cell types compared to the initial 4% in undifferentiated ES cells (Fig. 3a). The maternal mtDNA was undetectable during *in vivo* and *in vitro* differentiation of the non-reversed sibling cell line 3243ST-ES2 (Supplementary Table 3). By contrast, maternal mtDNA levels increased in all tested differentiated tissues in the

reversed ST-ES7 and NT-ES8 (Fig. 3a and Supplementary Tables 3, 4). These results demonstrate that the mtDNA reversion phenomenon is not unique to ES cells and occurs during differentiation, raising the possibility that reversal to the mutant mtDNA may occur in some MRT children.

Based on our observations that specific haplotype combinations led to a reversal of the maternal mtDNA, two mechanisms were

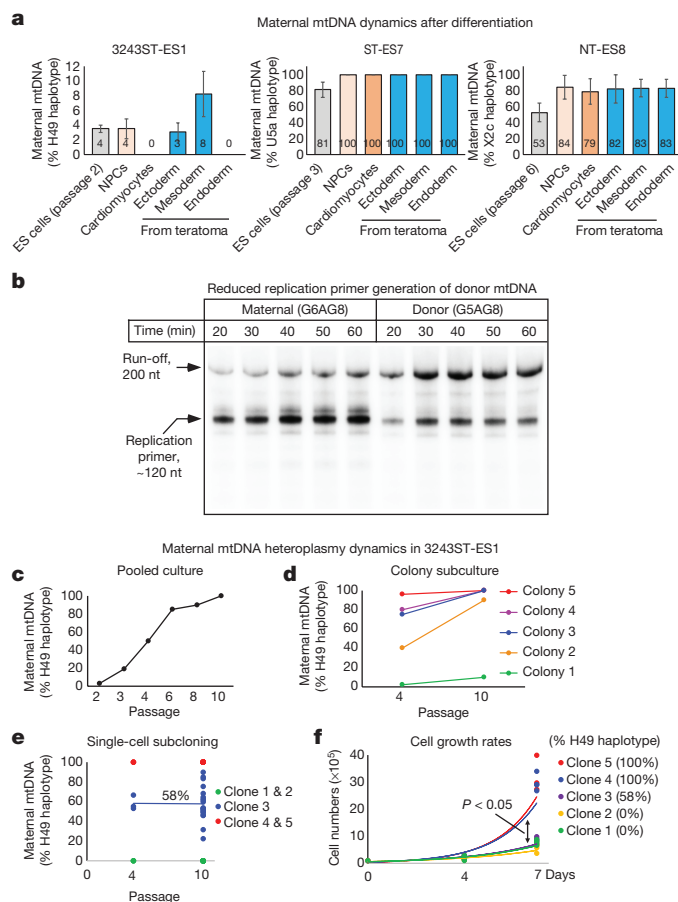


Figure 3 | Mechanisms affecting mtDNA reversal. **a**, Maternal mtDNA heteroplasmy changes in reversed 3243ST-ES1, ST-ES7 and NT-ES8 during *in vitro* and *in vivo* differentiation ($n = 1$). Bars represent a mean heteroplasmy of maternal mtDNA based on all SNP positions. Data are shown as mean \pm s.d. See also 3243ST-ES1, ST-ES7 and NT-ES8 in Supplementary Tables 3 and 4. **b**, Polymorphism in the CSBII region results in decreased efficiency of replication primer generation by mitochondrial RNA polymerase. The 311–315CCCC polymorphism is a shortening of the first G stretch in the G quadruplex region of mtDNA (G5 versus G6), responsible for transcription termination and replication primer generation. **c**, Heteroplasmy rates for the maternal mtDNA (H49 haplotype) gradually increased during extended *in vitro* culture of 3243ST-ES1 and reached homoplasmy ($n = 1$). See also Supplementary Table 3. **d**, The heteroplasmy levels also increased during ES cell colony subculture ($n = 1$). See also Supplementary Table 3. **e**, Maintenance of stable heteroplasmy in progeny of individually cultured cells. Heteroplasmy levels for maternal mtDNA were measured by amplification refractory mutation system–qPCR ($n = 3$ technical replicates) and whole mtDNA sequencing ($n = 1$). Data points at passage 10 are maternal mtDNA at individual SNPs. See also Supplementary Table 3. **f**, Faster cell growth and proliferation rates in clones containing higher maternal mtDNA heteroplasmy. $P < 0.05$; one-way ANOVA; $n = 4$ technical replicates. Source data files are available for **e** and **f**.

investigated: preferential replication of specific haplotypes and the capability of certain mtDNA endow cells with growth advantages^{24,25}. We first focused on the highly polymorphic D-loop region, called conserved sequence box II (CSBII)²⁵. It has been shown that a polymorphism in CSBII (G5AG7) affects efficiency of mitochondrial transcription termination and replication primer generation²⁵. When we analysed the donor and maternal mtDNA in a total of 26 ES cell lines comprised of 18 different haplotype combinations, we noted that two reversed sibling lines, ST-ES7 and ST-ES8 carried donor mtDNA with a G5AG8 polymorphism while the maternal haplotype was G6AG8 (Supplementary Table 5). Using *in vitro* transcription assays, we

analysed whether synthesis of the replication primer by mitochondrial RNA polymerase was affected and found that the deletion of a single guanosine residue (G5AG8 versus G6AG8) in donor mtDNA resulted in a fourfold reduction of replication primer synthesis (Fig. 3b). These results suggest that certain mtDNA haplotypes varying in the CSBII sequence provide more efficient synthesis of the replication primer that may, subsequently, confer replicative advantage. Sequence analysis did not find CSBII SNP differences in the remaining two reversed lines, 3243ST-ES1 and NT-ES8. However, a number of other D-loop polymorphisms were observed in the coreTAS region, which is also implicated in the regulation of mtDNA replication²⁶. Whereas the exact mechanisms by which these polymorphisms affect replication remain unclear, we speculate that, similar to the CSBII, they may contribute to replication bias of a particular mtDNA haplotype.

Because other replication-independent mechanisms may affect reversal, we also examined 3243ST-ES1 and NT-ES8 in whole cultures versus individual cell clones. Starting maternal mtDNA levels increased with passaging in whole cultures (Fig. 3c, d, Extended Data Fig. 7a and Supplementary Tables 3, 4). However, when we dissociated ES cells into individual cells and cultured isolated single-cell clones, the starting mtDNA levels varied in different clones but heteroplasmy did not change with passaging²³ (Fig. 3e and Supplementary Table 3). Notably, we observed substantial differences in cell proliferation and growth rates among different clones. Those with higher maternal mtDNA levels exhibited significantly faster growth rates ($P < 0.05$) than those with lower maternal mtDNA (Fig. 3f). These results suggest that certain mtDNA haplotypes confer ES cells with faster growth and proliferative advantages. Thus, in mixed cultures consisting of cells with varying degrees of heteroplasmy, a sub-population of cells with higher maternal mtDNA may overgrow cells with donor mtDNA. This possibility is independent of detectable mitochondrial activity, as complex I and complex IV enzyme activities were similar (Extended Data Fig. 7b).

In summary, among factors that might contribute to mtDNA reversal, several appear to be unlikely, including the MRT method (spindle transfer, pronuclear transfer or somatic cell nuclear transfer)^{27,28}, the presence of pathogenic mtDNA mutations and the genetic distance between donor and maternal mtDNA measured in total SNPs²⁸. A possible mechanism could be bias towards preferential amplification of mtDNA haplotypes with specific D-loop polymorphisms. Therefore, we propose selecting compatible donors harbouring CSBII or other D-loop sequences similar to the maternal mtDNA. While the number of tested combinations in our study was small, a proposed order of replication advantage in mtDNA is: H56>H1b, U5a>H1b>F1a, U5a>X2c>D4a and H49>B2k. However, stochastic or bottleneck mtDNA amplification during early embryo development may also coexist^{29,30}.

We show that some mtDNA haplotypes could affect cell growth and proliferation thereby providing selective advantage for cells with maternal mtDNA. Future applications of MRT will require additional studies evaluating compatible donor mtDNA haplotypes to avoid reversion to mutant mtDNA.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 March; accepted 2 November 2016.

Published online 30 November 2016.

- Archer, S. L. Mitochondrial dynamics—mitochondrial fission and fusion in human diseases. *N. Engl. J. Med.* **369**, 2236–2251 (2013).
- Koopman, W. J., Willems, P. H. & Smeitink, J. A. Monogenic mitochondrial disorders. *N. Engl. J. Med.* **366**, 1132–1141 (2012).
- Schon, E. A., DiMauro, S. & Hirano, M. Human mitochondrial DNA: roles of inherited and somatic mutations. *Nat. Rev. Genet.* **13**, 878–890 (2012).
- Wallace, D. C. & Chalkia, D. Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. *Cold Spring Harb. Perspect. Biol.* **5**, a021220 (2013).
- Gorman, G. S. *et al.* Mitochondrial donation—how many women could benefit? *N. Engl. J. Med.* **372**, 885–887 (2015).

6. Wolf, D. P., Mitalipov, N. & Mitalipov, S. Mitochondrial replacement therapy in reproductive medicine. *Trends Mol. Med.* **21**, 68–76 (2015).
7. Craven, L. *et al.* Pronuclear transfer in human embryos to prevent transmission of mitochondrial DNA disease. *Nature* **465**, 82–85 (2010).
8. Tachibana, M. *et al.* Mitochondrial gene replacement in primate offspring and embryonic stem cells. *Nature* **461**, 367–372 (2009).
9. Wallace, D. C. A mitochondrial bioenergetic etiology of disease. *J. Clin. Invest.* **123**, 1405–1412 (2013).
10. McClelland, K. *MELAS Within a family: Mitochondrial DNA Heteroplasm and Clinical Variability*. MSc thesis, California State University (2014).
11. Monnot, S. *et al.* Segregation of mtDNA throughout human embryofetal development: m.3243A>G as a model system. *Hum. Mutat.* **32**, 116–125 (2011).
12. Smeets, H. J., Sallevelt, S. C., Dreesen, J. C., de Die-Smulders, C. E. & de Co, I. F. Preventing the transmission of mitochondrial DNA disorders using prenatal or preimplantation genetic diagnosis. *Ann. NY Acad. Sci.* **1350**, 29–36 (2015).
13. Kelsey, T. W., Wright, P., Nelson, S. M., Anderson, R. A. & Wallace, W. H. A validated model of serum anti-Müllerian hormone from conception to menopause. *PLoS One* **6**, e22024 (2011).
14. Kelsey, T. W., Anderson, R. A., Wright, P., Nelson, S. M. & Wallace, W. H. Data-driven assessment of the human ovarian reserve. *Mol. Hum. Reprod.* **18**, 79–87 (2012).
15. Brown, D. T., Samuels, D. C., Michael, E. M., Turnbull, D. M. & Chinnery, P. F. Random genetic drift determines the level of mutant mtDNA in human primary oocytes. *Am. J. Hum. Genet.* **68**, 533–536 (2001).
16. Boucrot, L. *et al.* Relationship between diminished ovarian reserve and mitochondrial biogenesis in cumulus cells. *Hum. Reprod.* **30**, 1653–1664 (2015).
17. Woodson, J. D. & Chory, J. Coordination of gene expression between organellar and nuclear genomes. *Nat. Rev. Genet.* **9**, 383–395 (2008).
18. Kearney, H. M., Thorland, E. C., Brown, K. K., Quintero-Rivera, F. & South, S. T. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet. Med.* **13**, 680–685 (2011).
19. Tachibana, M. *et al.* Towards germline gene therapy of inherited mitochondrial diseases. *Nature* **493**, 627–631 (2013).
20. Paull, D. *et al.* Nuclear genome transfer in human oocytes eliminates mitochondrial DNA variants. *Nature* **493**, 632–637 (2013).
21. Tachibana, M. *et al.* Human embryonic stem cells derived by somatic cell nuclear transfer. *Cell* **153**, 1228–1238 (2013).
22. Ma, H. *et al.* Metabolic rescue in pluripotent cells from patients with mtDNA disease. *Nature* **524**, 234–238 (2015).
23. Kang, E. *et al.* Age-related accumulation of somatic mitochondrial DNA mutations in adult-derived human iPSCs. *Cell Stem Cell* **18**, 625–636 (2016).
24. Burgstaller, J. P., Johnston, I. G. & Poulton, J. Mitochondrial DNA disease and developmental implications for reproductive strategies. *Mol. Hum. Reprod.* **21**, 11–22 (2015).
25. Agaronyan, K., Morozov, Y. I., Anikin, M. & Temiakov, D. Mitochondrial biology. Replication–transcription switch in human mitochondria. *Science* **347**, 548–551 (2015).
26. Jemt, E. *et al.* Regulation of DNA replication at the end of the mitochondrial D-loop involves the helicase TWINKLE and a conserved sequence element. *Nucleic Acids Res.* **43**, 9262–9275 (2015).
27. Yamada, M. *et al.* Genetic drift can compromise mitochondrial replacement by nuclear transfer in human oocytes. *Cell Stem Cell* **18**, 749–754 (2016).
28. Hyslop, L. A. *et al.* Towards clinical application of pronuclear transfer to prevent mitochondrial DNA disease. *Nature* **534**, 383–386 (2016).
29. Lee, H. S. *et al.* Rapid mitochondrial DNA segregation in primate preimplantation embryos precedes somatic and germline bottleneck. *Cell Reports* **1**, 506–515 (2012).
30. Koehler, C. M. *et al.* Replacement of bovine mitochondrial DNA by a sequence variant within one generation. *Genetics* **129**, 247–255 (1991).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors acknowledge the OHSU Embryonic Stem Cell Research Oversight Committee and the Institutional Review Board for oversight and guidance. We thank all study participants for tissue donations and the Women's Health Research Unit staff, University Fertility Consultants and the Reproductive Endocrinology and Infertility Division in the Department of Obstetrics and Gynecology, Oregon Health and Science University for support and procurement of human gametes. Studies were supported by the Leduq Foundation, OHSU institutional funds and Cincinnati Children's Hospital Research Foundation. Work in the laboratory of J.C.I.B. was supported by the G. Harold and Leila Y. Mathers Charitable Foundation, the Leona M. and Harry B. Helmsley Charitable Trust and the Moxie Foundation. A.P.L. was partially supported by a fellowship from the Hewitt Foundation. P.M.R. was partially supported by a fellowship from Fundación Alfonso Martín Escudero.

Author Contributions E.K., N.M.G., P.A. and S.M. conceived the study and designed the experiments. T.Hu., P.A. and A.K. evaluated clinical genetic results and coordinated recruitment of study participants. P.A. and J.J. performed skin biopsies and blood collections. P.A., D.B., D.L. and D.W. performed ovarian stimulations and oocyte retrievals. N.M.G., R.T.H. and S.M. conducted spindle transfer, intracytoplasmic sperm injection, embryo culture and establishment of ES cell lines. E.K., R.T.H., H.M., Y.Le., Y.Li, R.A., T.Ha. and N.M.G. cultured ES cells. E.K., C.V.D., Y.Le. and T.Ha. performed teratoma studies. E.K., C.V.D., Y.Li and D.J. performed mitochondrial complex assays. J.W., P.M.R. and A.P.L. performed *in vitro* differentiation and Seahorse assays. K.A. and D.T. performed mtDNA transcription experiments. E.K., C.V.D., R.T.H. and A.K. prepared mtDNA and performed MiSeq assays. E.K. and Y.Le. performed NGS data analysis and interpretation. E.K., C.V.D. validated mtDNA mutations by Sanger sequencing. H.M. and R.T.H. performed ARMS-qPCR. X.W., S.L. and T.Hu. performed CNV data analysis. R.K. and C.C. performed the comparative genome hybridization array. S.O. performed G-banding karyotypes. E.K., J.W., D.P.W., J.C.I.B., P.A. and S.M. analysed data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.M. (mitalipo@ohsu.edu) and P.A. (amatop@ohsu.edu).

Reviewer Information *Nature* thanks E. Shoubridge and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Study oversight. Initial study protocol approval was granted in 2011 by the Oregon Health & Science University (OHSU) Institutional Review Board (IRB) and strictly followed the guidelines set forth by the Oregon Stem Cell Research Oversight Committee (OSCR). OSCRO guidelines permit recruitment of research gamete donors and subsequent fertilization and generation of pre-implantation embryos for research purposes. The protocol allows testing various MRT approaches prior to and after fertilization and pre-implantation. MRT embryo development can be terminated or used for ES cell derivation as a critical step in testing the safety and efficacy of novel cell and gene therapy treatments before clinical applications. The study protocol has undergone annual continuing reviews and extensions by IRB which includes biannual monitoring of study-related procedures involving human gametes, embryos and ES cells. No federal funds were used for this study in accordance with federal policies.

Participants. Healthy oocyte and sperm donors were recruited locally, via print- and web-based advertising, through OHSU Women's Health Research Unit. Oocyte donors were required to meet strict research inclusion criteria to minimize participant risk. Prior to consenting, eligible participants attended an information session, describing study goals and related clinical procedures. Carriers of mtDNA mutations were identified nationwide through print- and web-based advertising (UMDF and OHSU websites). Five families with existing children diagnosed with mitochondrial disease through MitoClinic in Cincinnati Children's Hospital Medical Center and OHSU Doernbecher Children's Hospital were identified and recruited. All families were out of state and travelled with children to OHSU for all study-related screening and procedures. Carrier oocyte donors underwent additional screening for study inclusion; strict inclusion criteria were met to minimize participant risk. Additionally, children from carrier donors provided samples for confirmation of pathogenic transmission of mtDNA mutations.

Informed consent. Written informed consent was obtained from gamete donors before any study-related procedures after a full review of study goals and clinical procedures. Consent forms for all gamete donors state that oocytes will be fertilized with sperm to generate pre-implantation embryos strictly for research purposes to develop and test novel MRT approaches. Consent forms also state that donated gametes and embryos eventually will be used for *in vitro* analyses or generation of ES cells but not for reproductive purposes. Written informed consent was obtained from parent before skin and blood donations from their underage children. As part of the consenting process, all participants provided written acknowledgement of the potential for incidental findings. If the participant requested to be informed, they were provided with some research findings by a physician collaborator and appropriate referrals were made. Patients undergoing standard IVF treatments were also contacted to donate immature oocytes that are routinely discarded as part of the IVF process. These participants were contacted before their scheduled oocyte retrievals. Written informed consent was obtained before egg retrieval.

Compensation. Research gamete donors were compensated for their time, effort, discomfort and inconvenience associated with the donation process at rates similar to gamete donors for reproductive purposes. Travel and lodging reimbursements were provided for mtDNA families travelling from out of state.

Controlled ovarian stimulation. Baseline screening was conducted including a medical history and physical exam, assessment of body mass index, anti-Müllerian hormone level and antral follicle count (AFC). Subjects underwent ovarian stimulation and oocyte retrieval employing standard IVF protocols and procedures¹⁹. All study-related procedures for oocyte donation took place at OHSU.

Spindle transfer. Spindle transfer was performed as previously described¹⁹. Briefly, meiotic metaphase II (MII) spindles were visualized under polarized microscopy and isolated with minimal cytoplasm (karyoplast, maternal mtDNA). The karyoplast was then placed in the perivitelline space of an enucleated oocyte (cytoplast, donor mtDNA) and fused using HVJ-E (haemagglutinating virus of Japan-envelope).

Oocyte vitrification. Oocyte vitrification was performed as previously described using commercially available human oocyte vitrification and thawing kits (Vitrification kit and Vitrification warming kit, Life Global).

Derivation of ES cells and culture. Spindle transfer blastocysts were freed from their zona pellucida and plated on mouse embryonic fibroblast (MEF) feeder layers on 4-well culture dishes for 5–7 days at 37°C, 3% CO₂, 5% O₂ and 92% N₂ in medium (DMEM/F12 with 10% fetal bovine serum (FBS), 10% knockout serum replacement (KSR), 0.1 mM non-essential amino acids, 1 mM L-glutamine, 0.1 mM β-mercaptoethanol, 5 ng ml⁻¹ basic fibroblast growth factor (bFGF), 10 μM ROCK inhibitor) as previously described¹⁹. Attached blastocyst outgrowths were manually minced into small pieces, re-plated onto fresh plates and maintained in knockout DMEM medium (Invitrogen) supplemented with 20% KSR, 0.1 mM non-essential amino acids, 1 mM L-glutamine, 0.1 mM β-mercaptoethanol, penicillin–streptomycin and 4 ng ml⁻¹ bFGF for further propagation and analysis. All cell lines were negative for mycoplasma contamination.

mtDNA transcription assays. *In vitro* transcription anti-termination assays were performed as described previously²⁵ using PCR-amplified templates containing region 202–481 of donor (H1b) or maternal (U5a) mtDNA. The products of the transcription reactions were resolved by 20% PAGE containing 6 M urea and visualized by PhosphorImager (GE Health).

Cell growth and proliferation assays. ES cells were adapted to grow under feeder-free conditions on Matrigel matrix in mTeSR1 medium (STEMCELL technologies)³¹. The cells were dissociated with Accutase (Life Technologies) for 2 min and approximately 10⁵ cells were seeded into each 60-mm dish. The cells were periodically harvested and counted using a Countess Automated Cell Counter (Invitrogen).

NPC differentiation and culture. A previously published protocol with minor modifications was used for NPC differentiation²². In brief, ES cells were maintained on MEFs in CDF12 medium before NPC differentiation. CDF12 medium contained DMEM/F12 (Life Technology), 20% KSR (Life Technologies), 2 mM Glutamax (Life Technologies), 0.1 mM non-essential amino acids (Life Technology), 0.1 mM β-mercaptoethanol (Life Technologies) and 4 ng ml⁻¹ FGF2 (Peprotech). ES cells were disaggregated with Collagenase IV (Life Technologies) and allowed to grow to about 40% confluence. To initiate neural induction, cells were washed twice with 1 × DPBS without calcium and magnesium (Corning Cellgro) and the medium was then changed to Neural Induction Medium 1 (NIM-1: 50% Advanced DMEM/F12 (Invitrogen), 50% Neurobasal (Invitrogen), 1 × B27 (Invitrogen), 1 × N2 (Invitrogen), 2 mM Glutamax and 10 ng ml⁻¹ hLIF (Peprotech), 4 μM CHIR99021 (Selleckchem), 3 μM SB431542 (Selleckchem), 2 μM dorsomorphin (Sigma) and 0.1 μM Compound E (EMD Chemicals Inc.)). After 2 days in culture in NIM-1 medium, cells were switched to Neural Induction Medium 2 (NIM-2: 50% Advanced DMEM/F12, 50% Neurobasal, 1 × N2, 1 × B27, 2 mM GlutaMAX and 10 ng ml⁻¹ hLIF, 4 μM CHIR99021, 3 μM SB431542 and 0.1 μM Compound E). Cells were further cultured in NIM-2 for 5 days with daily medium change. On the last day in NIM-2 medium, cells were treated overnight with 10 μM Y27632 (Selleckchem) and 20–30 'dome'-shaped colonies were manually picked and digested with Accumax (Innovative Cell Technologies) for 10 min at 37°C. After Accumax treatment cells were gently disaggregated into single cells and re-plated onto Matrigel-coated 6-well plates at a density of about 3.5 × 10⁴ per cm² in Neural Progenitor cell Maintenance Medium (NPM: 50% Advanced DMEM/F12, 50% Neurobasal, 1 × B27, 1 × N2, 2 mM GlutaMAX, 10 ng ml⁻¹ hLIF, 3 μM CHIR99021 and 2 μM SB431542) supplemented with 10 μM Y27632. NPCs were maintained on Matrigel-coated dishes in NPM. NPCs were passaged when they reached around 70% to 80% confluence using Accumax and replated at a density of 3 × 10⁴ per cm² with daily medium change. For the initial six passages, NPCs were pre-treated with 10 μM Y27632 overnight before and during cell splitting. The study was randomized, and the investigators were blinded to treatment allocations in collaborators' laboratories.

Cardiomyocyte differentiation. Cardiomyocyte differentiation was performed based on a GiWi (GSK3 inhibitor and Wnt inhibitor) protocol described previously³². Briefly, ES cells were grown on Matrigel-coated plates in mTeSR1 medium to 80–90% confluence before passaging with Accutase and then re-seeded at 0.5–1.5 million cells per well in 12-well plates in 1 ml of mTeSR1 plus 10 μM Y27632. On day 5, cells were incubated with 6–12 μM CHIR99021 (Selleckchem) for 16–24 h after which the medium was replaced with 2 ml of RPMI/B27 (without insulin) and cultured for two days. On day 8, 2 ml combined medium was prepared by mixing 1 ml of medium collected from the 12-well plate with 1 ml fresh RPMI/B27 (without insulin) medium. Medium on the cells was then replaced with 2 ml combined medium with 5 μM IWP2 (Tocris) and cells were cultured for 48 h. At day 10, 2 ml of fresh RPMI/B27 (without insulin) was added to each well of the 12-well plate. Starting from day 12, medium was changed every three days with 2 ml per well of RPMI/B27 (with insulin). Contracting cardiomyocytes were observed as early as day 17 from the initial passaging of ES cells.

Differentiated cell types were characterized by immunocytochemistry as described previously²².

Teratoma assay. ES cells were injected into the femoral region of 6-week-old, male SCID mice (Charles River). Mice with tumours were euthanized and teratomas were isolated, sectioned and histologically characterized for the presence of representative tissues as described previously²².

Assessment of mitochondrial function. Mitochondrial respiratory chain enzymatic activities (complex I and IV) were measured by spectrophotometry, as previously described^{23,33}. In brief, intact mitochondria were isolated from fibroblasts and treated with and without 3 mM rotenone at 340 nm wavelength for 5 min. Differences of absorbance per minute were obtained, and the specificity of COMI activity was estimated by the percentage of rotenone inhibition.

For live cell oxygen consumption, a XF96 Extracellular Flux Analyzer (Seahorse Biosciences) was used to measure oxygen consumption rates (OCR) and extracellular acidification rates as described previously²². In brief, NPCs were seeded at a density of 30,000 cells per well of a XF96 cell-culture microplate and incubated

for 24 h to allow cells to attach. Prior to the assay, NPCs were equilibrated for 1 h in unbuffered XF-assay medium supplemented with 25 mM glucose, 1 mM sodium pyruvate, 2 mM glutamax, 1 × non-essential amino acids and 1% (v/v) FBS in a non-CO₂ incubator. Sequential compound injections of oligomycin (0.5 µg ml⁻¹), carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone (FCCP, 1 µM), rotenone (0.5 µM) and antimycin A (1 µM) were used to measure the parameters of mitochondrial respiration. Indices of mitochondrial function were calculated as basal respiration rate (baseline OCR—rotenone or antimycin A OCR), ATP-dependent (basal respiration rate—oligomycin OCR), maximal respiration rate (FCCP OCR—rotenone or antimycin A OCR) and oxidative reserve (maximal respiration rate—basal respiration rate). Each plotted value was the mean of a minimum of four replicate wells, and was normalized to total cell numbers plated. Results were presented as means ± s.e.m. One-way ANOVA was used for three group comparisons and Student's *t*-test was used for pairwise comparisons. *P* < 0.05 was considered significant. The study was randomized, and the investigators were blinded to sample allocations among different groups.

mtDNA sequencing. Mutations and general mtDNA haplotypes were determined in oocytes, blood, skin fibroblasts, urine, ES cells and differentiated tissues by whole mtDNA next generation sequencing using the Illumina MiSeq platform^{22,23}. DNA was extracted from cells or tissues using Gentra DNA extraction kit (Qiagen), and from oocytes using Pico Pure DNA Extraction Kit (Life Technologies). mtDNA was amplified by a single PCR reaction as previously described²². mtDNA amplification from individual oocytes was performed using 2 primer sets: 7272F 5'-GGCTCATTCATTTCTCTAACAG-3', 15712R 5'-TTGGCTTAGTGGGCGAAATA-3' and 15635F 5'-TCCATCCTCATCC TAGCAAT-3', 7401R 5'-GGGGGCATCCATATAGTCAC-3'.

Detection of mtDNA variants by Sanger sequencing. mtDNA variants present at over 10% heteroplasmy were corroborated independently by Sanger sequencing. Regions of mtDNA containing germline mutations or SNPs were amplified by PCR with previously reported primer sets using PCR Platinum SuperMix High Fidelity kit (Life Technologies)²³. PCR products were purified, sequenced and analysed by Sequencer v5.0 (GeneCodes).

ARMS-qPCR and copy number measurement. The amplification refractory mutation system quantitative PCR assay (ARM-qPCR) was performed to verify

heteroplasmy at mt8993T>G, mt13513G>A and mt3243A>G as previously described²² and to measure carryover in control spindle transfer embryos and ES cells at 7843A>G, 16519T>C and 16278T>C¹⁹. Measurement of mtDNA copy number was performed as previously described³⁴.

Genetic and cytogenetic analyses. Karyotyping by G-banding was performed on 20 metaphase cells from each ES cell line by the OHSU Cytogenetics Laboratory. The comparative genome hybridization array was performed by IVI Gen. CNVs were identified by SNP genotyping as previously described³⁵ and analysed at the clinical genetic diagnosis laboratory in Cincinnati Children's Hospital Medical Center.

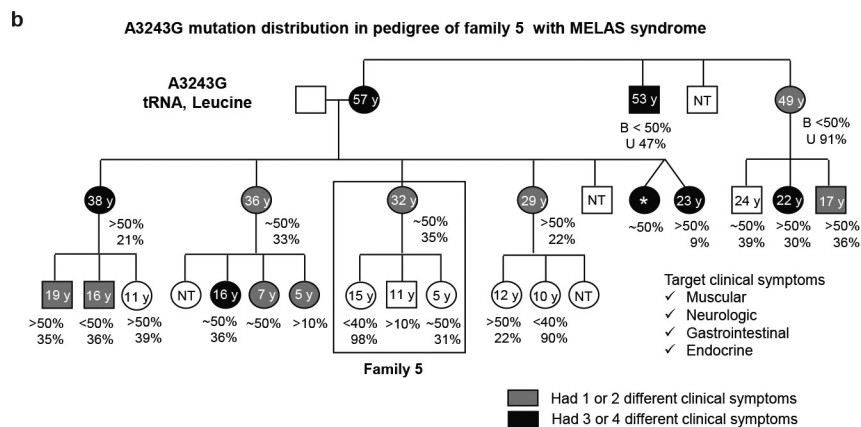
Statistical analysis. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment except for neural progenitor cell and cardiomyocyte differentiations and measurements of mitochondrial oxygen consumption rates. Results are presented as means ± s.d. or s.e.m. *P* < 0.05 was considered significant. Data were analysed using Pearson non-parametric test for correlation, independent group *t*-test or χ^2 test for pairwise comparison and ANOVA with Bonferroni analysis for multiple comparisons (IBM SPSS).

Data availability. All data generated or analysed during this study and included in this letter (and its Supplementary Information) are deposited in the NCBI SRA (SRP091883) and GEO (GSE87897).

31. Wu, J. *et al.* An alternative pluripotent state confers interspecies chimaeric competency. *Nature* **521**, 316–321 (2015).
32. Lian, X. *et al.* Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/ β -catenin signaling under fully defined conditions. *Nat. Protoc.* **8**, 162–175 (2013).
33. Spinazzi, M., Casarin, A., Pertegato, V., Salvati, L. & Angelini, C. Assessment of mitochondrial respiratory chain enzymatic activities on tissues and cultured cells. *Nat. Protoc.* **7**, 1235–1246 (2012).
34. Monnot, S. *et al.* Mutation dependence of the mitochondrial DNA copy number in the first stages of human embryogenesis. *Hum. Mol. Genet.* **22**, 1867–1872 (2013).
35. Ma, H. *et al.* Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature* **511**, 177–183 (2014).

a Information on families with mitochondrial disease

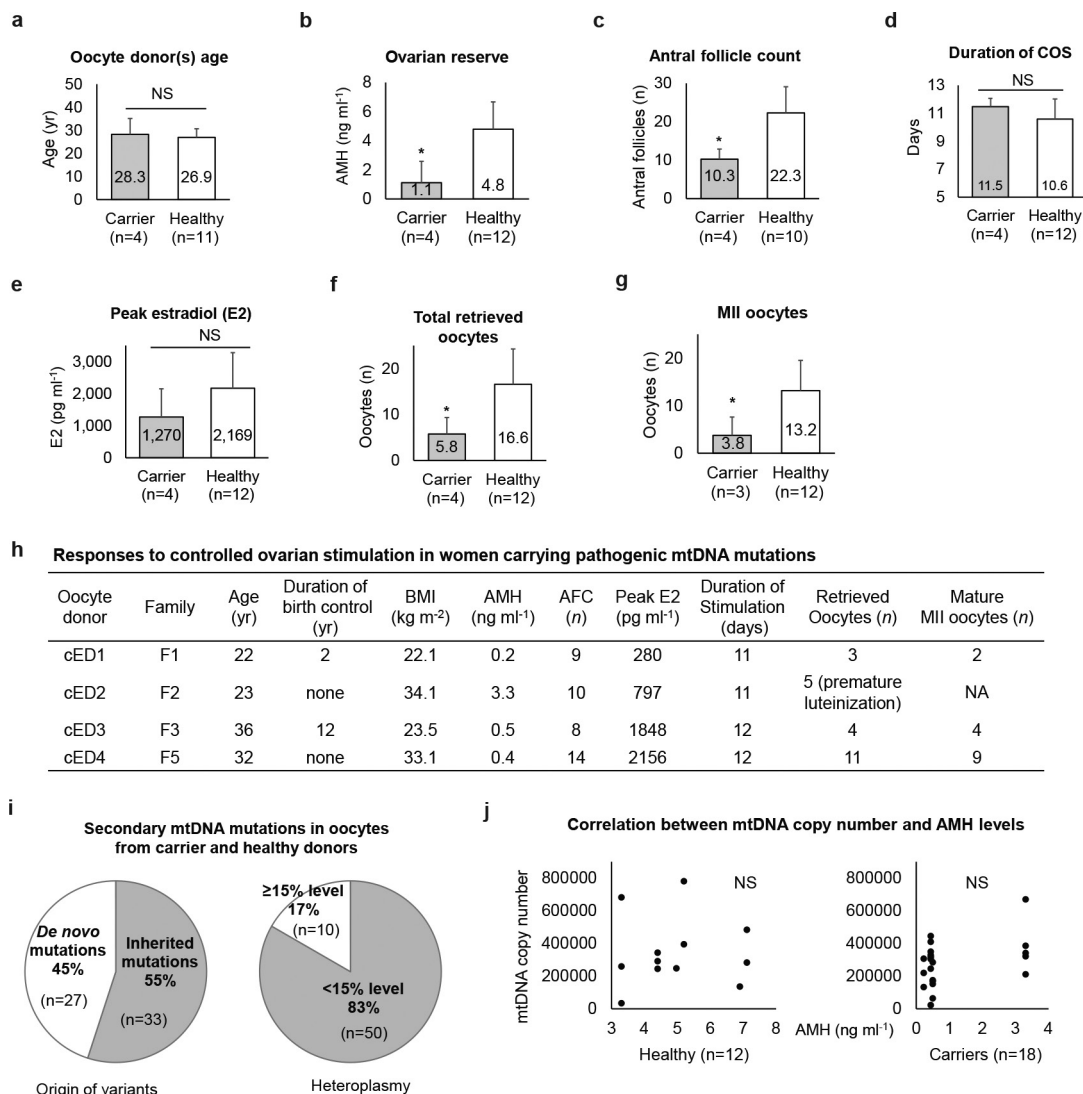
Mitochondrial disease	Family	Affected member	Age (yr)	Clinical manifestation	Mother's age (yr)	mtDNA mutation	Gene
Leigh syndrome	F1	Daughter	2	Seizures, metabolic acidosis, developmental delay, difficulty to thrive, abnormal basal ganglia and enlargement heart ventricle	22	T8993G	<i>ATP6</i> (non-syn)
	F2	Son	2.5	Seizures, chronic acidosis, developmental delay, difficulty to thrive and encephalopathy	23	T8993G	<i>ATP6</i> (non-syn)
	F3	Son	12	Seizures, failure to thrive, difficulty in breathing, enlarged heart ventricle, kidney failure and optic nerve atrophy	36	G13513A	<i>ND5</i> (non-syn)
	F4	Daughter	1	Seizures, difficulty to thrive, lactic acidosis, neuro-motor impairment and history of congenital brain abnormality	28	No	NA
MELAS syndrome	F5	Maternal sister	23	Mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes	32	A3243G	tRNA (Leucine)



Extended Data Figure 1 | Families with mitochondrial disease.

a, Families with mitochondrial disease and women of reproductive age were recruited for MRT. All Leigh syndrome families had an existing, severely affected child. F, family; non-syn, non-synonymous; NA, not applicable. **b**, Family 5 was selected from an extensive MELAS syndrome

pedigree. Clinical phenotypes of A3243G MELAS syndrome varied even with similar heteroplasmy levels. Asterisk indicates the first MELAS patient diagnosed clinically. Heteroplasmy in blood (upper panel) and urine (lower panel). Squares, males; circles, females. B, blood; U, urine. NT, not tested. Percentage values refer to heteroplasmy; y, years-old.



Extended Data Figure 2 | Response to controlled ovarian stimulation and oocyte characteristics. **a**, Age of oocyte donor(s) was similar between carrier and healthy groups. **b**, **c**, Ovarian reserve (**b**) and antral follicle count (**c**) were significantly lower in carriers than in healthy oocyte donors. **d**, **e**, Duration of controlled ovarian stimulation (**d**, COS) was longer in carriers and peak oestradiol (**e**, E2) on day of hCG tended to be lower in carriers. **f**, **g**, Total oocyte yield (**f**) and number of mature oocytes (**g**) were significantly lower in carriers than healthy oocyte donors. *n* = the number of controlled ovarian stimulation cycles. Data are represented as mean \pm s.d. **P* < 0.05. **h**, Baseline characteristics and cycle outcomes.

Carriers had a lower anti-Müllerian hormone (AMH) levels. Lower peak oestradiol levels were measured in cED1 and cED2. Birth control method: combined oral contraceptive, medroxyprogesterone. **i**, Analysis of heteroplasmic mtDNA variants detected in MII oocytes. *De novo* indicates unique mutations found in individual oocytes; inherited indicates mutations shared with other oocytes, sibling children or egg donors. *n* = the number of mutations in individual oocytes. **j**, anti-Müllerian hormone levels, a measure of ovarian reserve, were not correlated with mtDNA copy number. *n* = the number of oocytes. NS, not significant. Source data files are available for **a–g**.

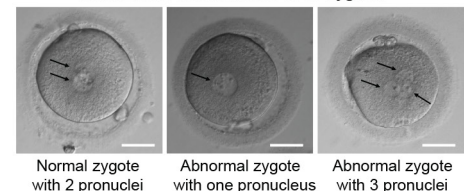
a

Donor oocyte mtDNA haplotypes and SNP differences in ST embryos

Cyto Karyo	Egg donor	hED1 (21-yr)	hED2 (29-yr)	hED3 (23-yr)	hED4 (25-yr)	hED5 (25-yr)	hED6 (28-yr)	hED7 (24-yr)	hED8 (34-yr)	hED9 (32-yr)	hED10 (27-yr)	hED11 (29-yr)
Egg donor	mtDNA haplotype	L2c	D4a	A2g	B2k	U5a	V3	T2	H1b	H2a	H6a	H56
hED1 (21-yr)	L2c									57		
hED2 (29-yr)	D4a			49								
hED3 (23-yr)	A2g		49									
hED5 (25-yr)	U5a						33		33			
hED6 (28-yr)	V3					33						
hED8 (34-yr)	H1b					33						
hED9 (32-yr)	H2a	57										6
hED11 (29-yr)	H56									6		
cED1 (22-yr)	H1b										21	
cED3 (36-yr)	T2b							22				
cED4 (32-yr)	H49				32							

b

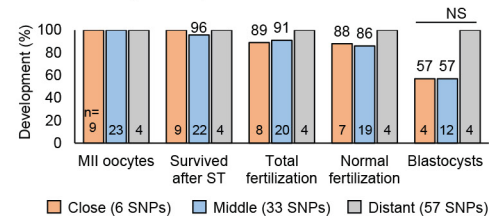
Fertilization abnormalities in ST zygotes



Normal zygote with 2 pronuclei Abnormal zygote with one pronucleus Abnormal zygote with 3 pronuclei

e

Embryo development with different donor mtDNA



c Fertilization abnormalities in ST zygotes and embryos

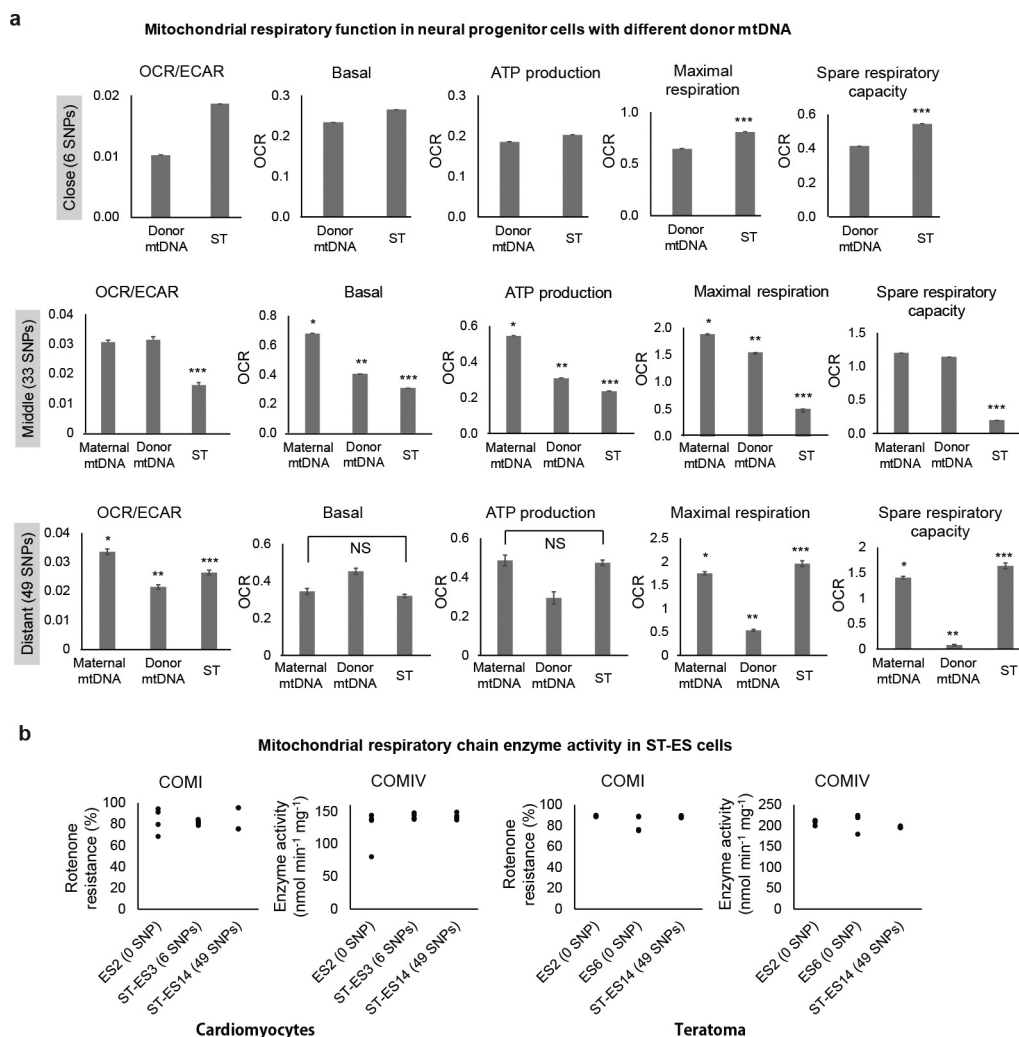
Treatment	Total fertilized zygotes (n)	Total abnormal fertilized zygotes (%)	Abnormal No. of pronuclei			Blastocysts	Karyotypes
			One	Three (%)	Abnormal divided		
Intact control	17	1 (6)	1	0	0	0	
Control ST	32	2 (6)	1	1 (3)	0	1	69,XXY
Carrier ST	12	4 (33)	1	2 (17)	1	0	

d Fertilization and embryo development of frozen ST oocytes

Treatment	MI I oocytes (n)	Survived after ST (%)	Total fertilization (%)	Normal fertilization (%)	Abnormal No. of pronuclei		Normal morula (%)	Normal blastocysts (%)
					One	Three (%)		
Vitrified cytoplasts/ Fresh spindles	6	6 (100)	6 (100)	4 (67)	1	1 (17)	4 (100)	1 (25)
Vitrified spindles/ Fresh cytoplasts	6	6 (100)	3 (50)	2 (67)	0	1 (33)	2 (100)	1 (50)

Extended Data Figure 3 | Fertilization and development after spindle transfer. **a**, Whole mtDNA sequencing was performed to identify haplotypes of all egg donors and spindle transfer was performed to match various haplotypes. Cyto, cytoplasm; Karyo, karyoplast. **b**, Normal fertilization (2 pronuclei) and abnormal fertilization (1 or 3 pronuclei); see arrows. Scale bars, 50 μ m. **c**, Carrier spindle transfer showed higher abnormal fertilization than controls. One zygote from the control

spindle transfer group with three pronuclei developed to a blastocyst and displayed a 69,XXY karyotype. **d**, The blastulation rate was similar between vitrified cytoplasm with fresh spindles and vitrified spindles with fresh cytoplasm ($P > 0.05$; t -test). **e**, Control spindle transfer embryo development as a function of donor mtDNA matching distances. The numbers on the top of bars are percentage of embryo development. n = the number of embryos; NS, not significant ($P > 0.05$; t -test).



Extended Data Figure 4 | Mitochondrial function in cells with donor mtDNA. **a**, Energy metabolism ($\frac{OCR}{ECAR}$) was measured and compared between neural progenitor cells (NPCs) derived from MRT and control embryos carrying either donor or maternal mtDNA. OCR, oxygen consumption rate (representing oxidative phosphorylation); ECAR, extracellular acidification rate (representing glycolysis). NPCs from MRT ES cells with 6 and 49 SNP differences displayed comparable oxygen consumption rates to NPCs harbouring maternal or donor mtDNA except maximal respiration in the 49 SNP group. The oxygen consumption rate was reduced in NPCs with 33 SNPs. The oxygen consumption rate data

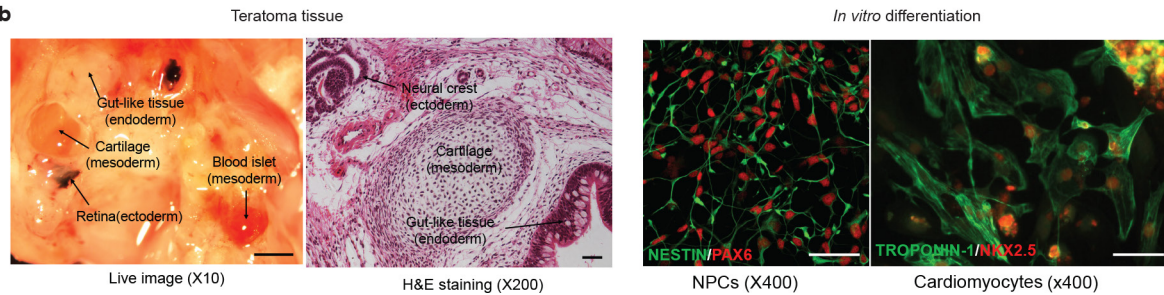
were normalized by live cell DNA content. * $P < 0.05$ versus host mtDNA, ** $P < 0.05$ versus donor mtDNA, *** $P < 0.05$ versus spindle transfer (t -test for 6 SNPs and one-way ANOVA for 33 and 49 SNPs, $n = 16$ technical replicates). Data are shown as mean \pm s.e.m. NS, not significant. **b**, Mitochondrial respiratory chain enzyme activity in differentiated cells from spindle transfer ES cell lines. Mitochondrial complex I and IV activities in differentiated cells from spindle transfer ES cell lines carrying donor mtDNA with 49 SNP differences were comparable to controls with either donor or maternal mtDNA ($P > 0.05$). One-way ANOVA, $n = 4$ technical replicates. Source data file is available for **b**.

a

***In vivo* and *in vitro* differentiation of ST-ES cells**

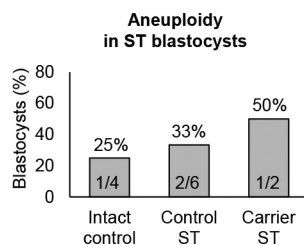
No.	Cell line	Maternal mtDNA haplotype	Donor mtDNA haplotype	SNP difference	Teratoma	NPCs	Cardiomyocytes
1	ES1	-----	H1b control -----	0	NT	yes	yes
2	ES2	-----	D4a control -----	0	yes	yes	yes
3	ES3	-----	H56 control -----	0	NT	yes	yes
4	ES6	-----	A2g control -----	0	yes	yes	yes
5	ST-ES3	H2a	H56	6	NT	yes	yes
6	ST-ES6	H1b	U5a	33	NT	yes	yes
7	ST-ES7	U5a	H1b/ U5a (reversed)	33	NT	yes	yes
8	ST-ES14	D4a	A2g	49	yes	yes	yes
9	ST3243-ES1	H49	B2k/ H49 (reversed)	31	yes	yes	yes
10	ST3243-ES2	H49	B2k	31	yes	yes	yes
11	NT-ES8	X2c	D4a/ X2c (reversed)	39	yes	yes	yes

b



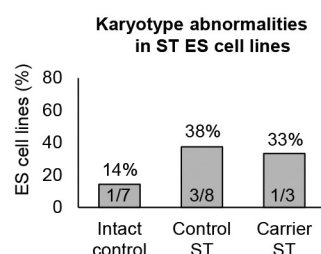
Extended Data Figure 5 | Differentiation of ES cells derived from MRT embryos. **a**, Summary of *in vitro* and *in vivo* differentiation of ES cells derived from spindle transfer or SCNT embryos carrying donor mtDNA. Similar to control IVF ES cells, all tested MRT ES cells produced teratoma tumours *in vivo* and formed neural progenitor cells (NPCs)

and cardiomyocytes *in vitro*. NT, not tested. **b**, Histological analysis of differentiated tissues from MRT ES cells. The representative tissues were collected and used for mtDNA carry-over analysis and measurements of mitochondrial function. Scale bars, 1 mm (live image) and 50 μ m (haematoxylin and eosin (H&E) and immunochemistry staining).

a

Summary of array CGH in intact and ST blastocysts

Treatment	#	Age of nuclear donor (yr)	Egg donor ID	aCGH Results	Sex
Intact control	1	34	hED8	Abnormal: -21	Male
	2	29	hED11	Normal	Male
	3	25	hED4	Normal	Female
	4	25	hED4	Normal	Male
Control ST	1	34	hED8	Abnormal: +7, +15	Male
	2	32	hED9	Normal	Male
	3	32	hED9	Normal	Female
	4	32	hED9	Normal	Male
	5	29	hED11	Abnormal: +12	Male
	6	21	hED1	Normal	Male
Carrier ST	1	36	cED3	Abnormal: +9	Male
	2	32	cED4	Normal	Female

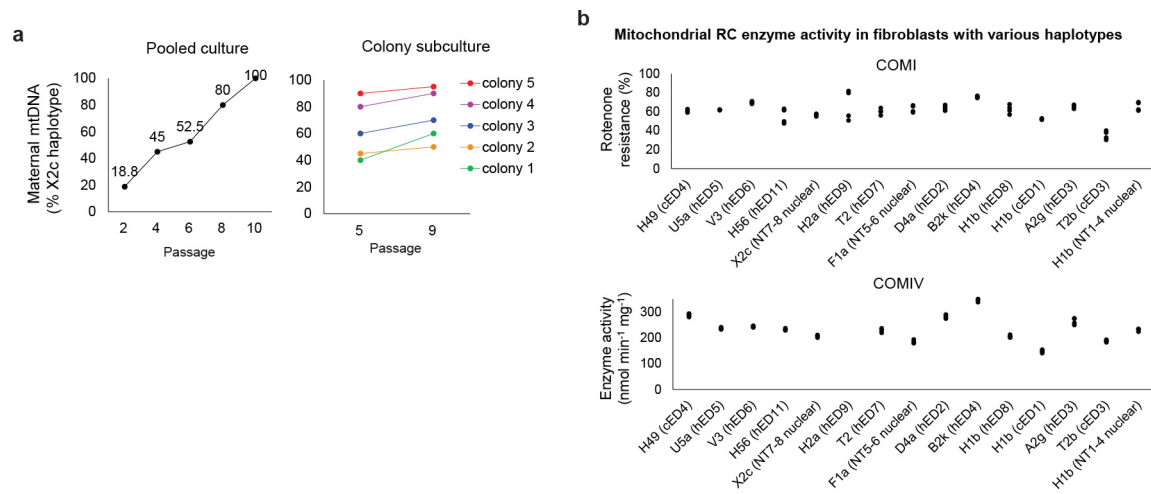
b

G-banding analysis of ES cell lines derived from intact and ST blastocysts

Treatment	#	Age of nuclear donor (yr)	Egg donor ID	ES cell line	Karyotype
Intact control	1	34	hED8	ES1	46,XX
	2	29	hED2	ES2	46,XX
	3	29	hED11	ES3	46,XY
	4	25	hED4	ES4	46,XY
	5	25	hED5	ES5	46,XX
	6	23	hED3	ES6	46,XX
	7	23	hED3	ES7	45,XY,-21
Control ST	1	34	hED8	ST-ES5	Mosaicism 47,XX +10 [12]/46,XX [8]
	2	34	hED8	ST-ES6	46,XX
	3	32	hED9	ST-ES3	46,XX
	4	29	hED2	ST-ES14	46,XY
	5	28	hED6	ST-ES10	45,X
	6	25	hED5	ST-ES7	46,XY
	7	25	hED5	ST-ES8	92,XXXX
	8	25	hED5	ST-ES9	46,XY
Carrier ST	1	36	cED3	13513 ST-ES	47,XY,+9
	2	32	cED4	3243 ST-ES1	46,XX
	3	32	cED4	3243 ST-ES2	46,XY

Extended Data Figure 6 | Nuclear genome abnormalities associated with spindle transfer. **a**, Aneuploidy rate in blastocysts determined by a comparative genome hybridization array was not significantly different in spindle transfer groups compared to controls. **b**, Karyotype abnormality

rate in ES cells determined by G-banding analysis was also comparable among spindle transfer groups ($P > 0.05$). Normal, 46 XX or 46 XY. Number inside bars, the number of blastocysts or ES cell lines (abnormal/total). Simple χ^2 tests.



Extended Data Figure 7 | Maternal mtDNA heteroplasmy changes and mitochondrial function. **a**, Maternal mtDNA (X2c haplotype) heteroplasmy increased during extended *in vitro* culture and reached homoplasmy in pooled cultures or in individual colony subcultures ($n = 1$). See also Supplementary Table 4. **b**, Respiratory chain complex 1

(COM I) and complex 4 (COM IV) enzyme activities were measured in fibroblasts carrying various human mtDNA haplotypes used in this study. No significant differences were noted ($P > 0.05$). One-way ANOVA, $n = 4$ technical replicates. Source data file is available for **b**.

Extended Data Table 1 | *De novo* copy number variations (CNVs) in ES cells derived from control and spindle transfer blastocysts

Treatment	#	ES cell line ID	CNVs	Chromosome Location	CNV chromosome	Interpretation
Intact control	1	ES1	Deletion (neutral)	15q11.2	male	699 kb region of chromosome 15 (15q11.2) shows low level mosaicism (approximately 10%-20%) for homozygosity of uncertain clinical significance.
	2	ES5	Deletion (neutral)	15q11.2	male	1.1 Mb region of chromosome 15 (15q11.2) shows mosaicism for homozygosity of uncertain clinical significance.
			Deletion (neutral)	Xp22.33-q28	male	Less than 10% of cells of the entire chromosome X (Xp22.33-Xq28) of uncertain clinical significance.
Control ST	1	ST-ES5	Duplication	10p15.3-q26.3	male	Mosaic trisomy 10.
			Deletion (neutral)	15q11.2	male	549 kb region of chromosome 15 (15q11.2) shows mosaicism for homozygosity of uncertain clinical significance.
	2	ST-ES6	Deletion	Xp22.33-q28	male	Monosomy X (Xp22.33-Xq28) of uncertain clinical significance.
			None	NA		NA
			None	NA		NA
			None	NA		NA
Carrier ST	1	3243ST-ES1	None	NA		NA
	2	3243ST-ES2	Deletion (neutral)	15q11.2	male	791 kb region of chromosome 15 (15q11.2) shows mosaicism for homozygosity of uncertain clinical significance.

Normal, diploid chromosomal arrangements are identical to original male and female parents. Grey box, shared CNVs among independent cell lines; NA, not applicable.

Extended Data Table 2 | MRT outcomes for families with pathogenic mtDNA mutations

	Family 1	Family 2	Family 3		Family 4		Family 5		
Mitochondrial disease	Leigh syndrome	Leigh syndrome	Leigh syndrome		Leigh syndrome		MELAS		
Pathogenic mtDNA mutation	T8993G	T8993G	G13513A		No		A3243G		
Age of carrier (yr)	22	23	36		28		32		
COS	Yes	Yes	Yes		Excluded		Yes		
No. retrieved oocytes	3	5	4		NA		11		
No. ST oocytes	2	Canceled	4		NA		7		
ST blastocysts	0	NA	2		NA		4		
Grade at D6	NA	NA	5AA	5BB	NA	5AA	5AA	5BB	5CC
Aneuploidy	NA	NA	Yes:+9	NT	NA	NT	No	No	No
Sex	NA	NA	M	NA	NA	NA	M	F	F
mtDNA carryover	NA	NA	< 1%	NA	NA	NA	< 1%	< 1%	< 1%*

Six spindle transfer embryos derived from carrier oocytes reached the blastocyst stage. At least two blastocysts were eligible for transfer. M, male; F, female. NA, not applicable; NT, not tested; D6, blastocysts at day 6 of development. The first number in grade indicates expansion, grade 5 means blastocyst is hatched. The second letter indicates quality of the inner cell mass (ICM), A refers to large ICM of tightly packed cells; B, smaller ICM, loose cells; C, very small ICM. The third letter grades the trophectoderm quality.

*mtDNA in ES cells from this blastocyst reversed to the original maternal haplotype but did not carry the A3243G mutation.

Intronic polyadenylation of PDGFR α in resident stem cells attenuates muscle fibrosis

Alisa A. Mueller^{1,2,3†}, Cindy T. van Velthoven^{1,2}, Kathryn D. Fukumoto^{1,2}, Tom H. Cheung^{1,2†} & Thomas A. Rando^{1,2,4}

Platelet-derived growth factor receptor α (PDGFR α) exhibits divergent effects in skeletal muscle. At physiological levels, signalling through this receptor promotes muscle development in growing embryos and angiogenesis in regenerating adult muscle^{1,2}. However, both increased PDGF ligand abundance and enhanced PDGFR α pathway activity cause pathological fibrosis^{3,4}. This excessive collagen deposition, which is seen in aged and diseased muscle^{5–7}, interferes with muscle function and limits the effectiveness of gene- and cell-based therapies for muscle disorders^{8,9}. Although compelling evidence exists for the role of PDGFR α in fibrosis, little is known about the cells through which this pathway acts. Here we show in mice that PDGFR α signalling regulates a population of muscle-resident fibro/adipogenic progenitors (FAPs) that play a supportive role in muscle regeneration but may also cause fibrosis when aberrantly regulated^{10–13}. We found that FAPs produce multiple transcriptional variants of *Pdgfra* with different polyadenylation sites, including an intronic variant that codes for a protein isoform containing a truncated kinase domain. This variant, upregulated during regeneration, acts as a decoy to inhibit PDGF signalling and to prevent FAP over-activation. Moreover, increasing the expression of this isoform limits fibrosis *in vivo* in mice, suggesting both biological relevance and therapeutic potential of modulating polyadenylation patterns in stem-cell populations.

Although aberrant PDGFR α signalling is linked to muscle pathology^{4,14–17}, we investigated how PDGFR α signalling is regulated during healthy muscle regeneration. PDGFR α is mostly unphosphorylated, and hence inactive, in resting muscle but is activated upon injury with levels peaking after 5 days (Extended Data Fig. 1a, b; for gel source data, see Supplementary Fig. 1). Accordingly, injured muscle shows a substantial increase of phospho-PDGFR α -positive cells (Extended Data Fig. 1c). Using a mouse containing a histone 2B (H2B) and enhanced green fluorescent protein (eGFP) fusion protein knocked into the *Pdgfra* locus², we found a population of PDGFR α -expressing cells located in interstitial regions (Extended Data Fig. 1d). Fluorescence-activated cell sorting (FACS) analysis revealed that PDGFR α is expressed strongly in over 99% of FAPs and that nearly all PDGFR α -expressing cells are from this population in both uninjured (Extended Data Fig. 1e–g) and injured (Extended Data Fig. 1h–m) muscle.

To understand how PDGFR α signalling is regulated in FAPs, we performed direct RNA sequencing because its single-molecule approach allows for straightforward quantification and also because it involves sequencing at the polyadenylated tail of RNA, providing the opportunity to evaluate post-transcriptional regulation via polyadenylation¹⁸. A previous study from our laboratory demonstrated that alternative polyadenylation at the 3' untranslated region (3' UTR) of transcripts has a critical role in modulating stem-cell behaviour¹⁹.

We found that FAPs produce not only transcriptional variants of *Pdgfra* that harbour different polyadenylation sites within the 3' UTR, but also multiple transcripts that are polyadenylated at introns within the protein-coding region (Fig. 1a). We confirmed polyadenylation at these sites through 3' rapid amplification of cDNA ends (RACE) (Fig. 1b and Extended Data Fig. 2a). Particularly noteworthy was a transcriptional variant (which we term 'In-PDGFR α ') that terminates within intron 16 of *Pdgfra*. This transcript, the most abundant of the intronic variants, contains the canonical AAUAAA polyadenylation termination motif preceded by an in-frame stop codon sequence, suggesting protein-coding potential. Such a protein would consist of a membrane-bound isoform of PDGFR α containing an intact ligand-binding domain but a truncated kinase domain (Extended Data Fig. 2b, c). This type of structural modification is notable because previous studies in cell lines have suggested that protein truncations produced through intronic polyadenylation can substantially affect functionality of the full-length counterpart²⁰. For example, artificial induction of intronic polyadenylation of tyrosine kinase receptors, in which both the kinase and transmembrane regions were truncated, caused cells to produce soluble decoy receptors that bound ligand or functional receptors, suppressing signalling²¹.

Given that the kinase domain of In-PDGFR α is disrupted, we proposed that this protein could act as a decoy to negatively regulate PDGF signalling. We therefore investigated whether the In-PDGFR α transcript produced a functional protein in FAPs. Indeed, whereas western blot analysis of FAP lysates using a C-terminal-directed antibody revealed a band at 180 kDa corresponding to the full-length protein (FL-PDGFR α) (Fig. 1c), an antibody directed at a central region of PDGFR α detected bands at both 180 kDa and 120 kDa (Fig. 1d). Overexpression of In-PDGFR α in FAPs using a retroviral construct resulted in increased levels of a 120-kDa protein (Fig. 1e).

Use of the identified stop codon in intron 16 would be expected to result in an extension of the stretch of 16 amino acids unique to In-PDGFR α (Extended Data Fig. 2d). We developed a polyclonal antibody directed against this specific 16-amino-acid peptide. This antibody detected a band at 120 kDa but not at 180 kDa in untreated FAPs. The 120 kDa band was also detected in cells overexpressing In-PDGFR α , demonstrating that it is specific for In-PDGFR α (Fig. 1e).

To investigate the specific regulation of In-PDGFR α production, we analysed FAPs activated in response to muscle injury. Overall levels of FL-PDGFR α increased but there was also a marked increase in the proportion of In-PDGFR α to FL-PDGFR α , peaking 3–5 days after injury (Fig. 1f), suggesting differential regulation of the intronic variant relative to the full-length form. After 3 weeks, the relative levels of In-PDGFR α were comparable to those in uninjured muscle (Fig. 1f). Using primers that could specifically detect In-PDGFR α and FL-PDGFR α (Extended Data Fig. 2d, e), we found a similar temporal pattern at the transcriptional level (Extended Data Fig. 2f).

¹Paul F. Glenn Center for the Biology of Aging, Stanford University School of Medicine, Stanford, California, USA. ²Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, California, USA. ³Program in Cancer Biology, Stanford University School of Medicine, Stanford, California, USA. ⁴Neurology Service and Rehabilitation Research and Development REAP, Veterans Affairs Palo Alto Health Care System, Palo Alto, California, USA. [†]Present addresses: Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA (A.A.M.); Division of Life Science, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China (T.H.C.).

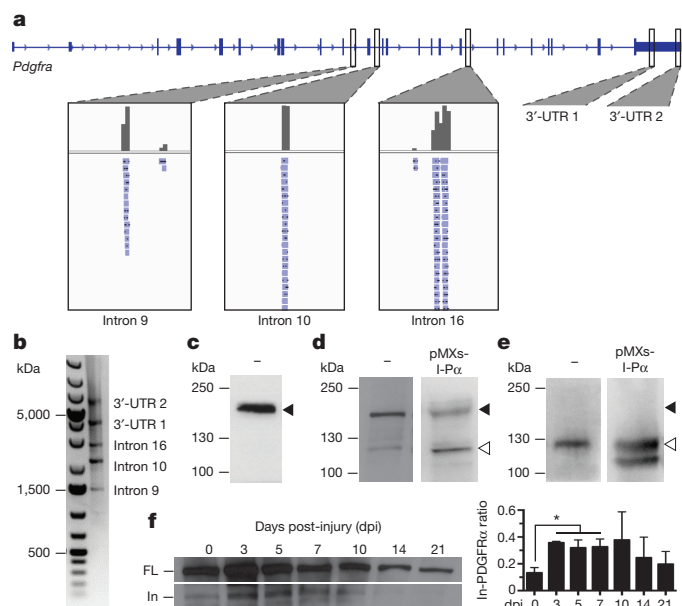


Figure 1 | *Pdgfra* undergoes intronic polyadenylation in FAPs, resulting in a truncated protein that increases in abundance during regeneration. **a**, Direct RNA sequencing reveals multiple polyadenylation sites for *Pdgfra*, including two sites within the 3' UTR and three sites within introns. **b**, 3' RACE, in which the gene-specific primer overlaps with the *Pdgfra* start codon, producing bands that correspond to the predicted sizes of the polyadenylation variant transcripts. **c–e**, Western blot analyses of FAPs which were untreated (–) or infected with an In-PDGFR α -expressing retrovirus (pMXs-I-P α). Lysates were probed with antibodies targeting the amino acid sequences at one of the following locations: the C terminus of FL-PDGFR α (**c**), the central region preceding exon 16 (**d**), and the C terminus unique to In-PDGFR α (**e**). Filled arrowheads denote 180 kDa; open arrowheads denote 120 kDa. Overexpression with In-PDGFR α resulted in a doublet (**e**); the lower band probably represents an alternative form of In-PDGFR α often seen for FL-PDGFR α (ref. 26). **f**, Levels of In-PDGFR α relative to FL-PDGFR α increase during regeneration. Lysates of FAPs collected from uninjured muscle (day 0) or at the indicated times after injury were probed by western blot with an antibody targeting a central region of PDGFR α . Expression level is plotted as the ratio of In-PDGFR α to FL-PDGFR α . $n = 3$ biological replicates of pooled FAPs per time point. Error bars represent s.e.m. * $P < 0.05$; unpaired Student's t -tests.

To evaluate the effect of intronic polyadenylation on PDGFR α activity, we used antisense morpholino oligonucleotides (AMOs), which can alter intronic polyadenylation site usage by blocking splicing and polyadenylation machinery²¹. We designed AMOs to target the intronic polyadenylation site (pA-AMOs) and the upstream 5' splice site (5'ss-AMO) expected to decrease and increase production of In-PDGFR α , respectively (Fig. 2a). We measured the effect of these morpholinos on both In-PDGFR α transcript and protein levels. The pA-AMOs decreased levels of In-PDGFR α without altering FL-PDGFR α expression (Fig. 2b, c and Extended Data Fig. 2g). Similarly, the 5'ss-AMO increased In-PDGFR α levels, and there was a corresponding reduction of FL-PDGFR α expression (Fig. 2d, e and Extended Data Fig. 2g).

We then investigated whether modulation of the levels of the intronic variant could alter PDGFR α signalling. Treatment of cells with PDGF-AA led to phosphorylation of PDGFR α and downstream components including ERK1/2 and Akt (Extended Data Fig. 3a, b). If In-PDGFR α serves as a decoy, we would expect decreased PDGFR α signalling when this variant is upregulated. As expected, pre-treatment of cells with the 5'ss-AMO, which increases In-PDGFR α abundance, led to reduced ERK1/2 and Akt phosphorylation (Fig. 2f, g), an effect also observed when we overexpressed In-PDGFR α (Extended Data

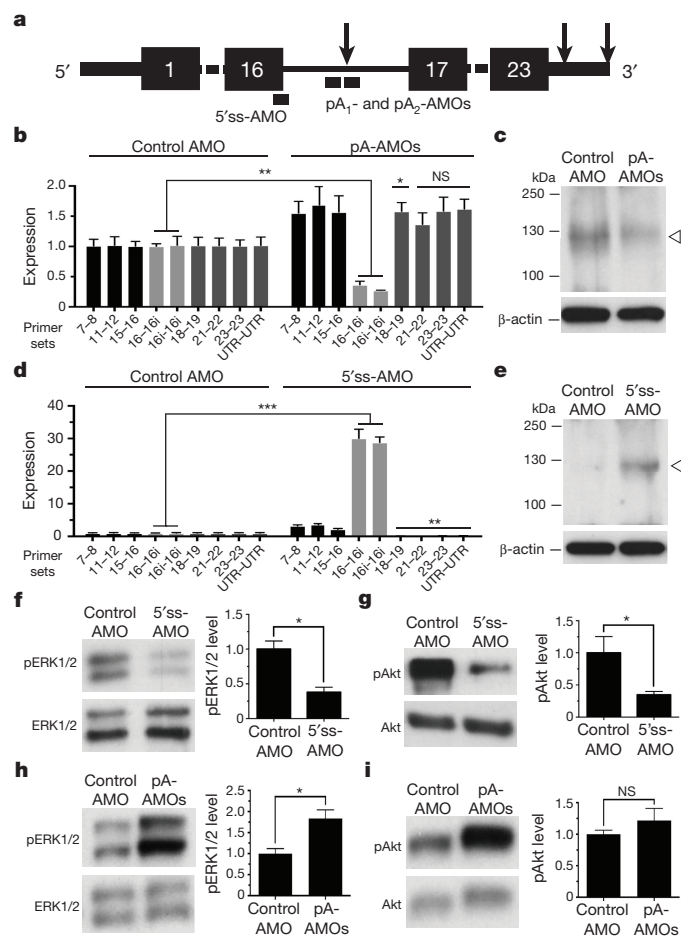


Figure 2 | Intronic polyadenylation of *Pdgfra* inhibits signalling through the receptor, whereas decreasing In-PDGFR α abundance augments signalling. **a**, Morpholino target location along *Pdgfra* transcript. Black boxes denote exon number. Arrows represent polyadenylation sites. **b–e**, Treatment of FAPs with pA-AMOs results in knockdown of the In-PDGFR α transcript (**b**) and protein (**c**) whereas treatment with the 5'ss-AMO upregulates the In-PDGFR α transcript (**d**) and protein (**e**). Open arrowheads denote 120 kDa in **c** and **e**. At the exposure shown in **e**, the expression of the endogenous In-PDGFR α protein is below the level of detection owing to the abundance of the protein in the 5'ss-AMO lane. **f**, **g**, Treatment of FAPs with the 5'ss-AMO decreases phosphorylation of ERK1/2 (**f**) and Akt (**g**) in response to PDGF-AA. **h**, **i**, Conversely, treatment of FAPs with the pA-AMOs increases phosphorylation of ERK1/2 (**h**) with a trend towards enhanced Akt phosphorylation (**i**) in response to PDGF-AA. pERK1/2 and pAkt levels represent the ratio of phosphorylated protein to total protein. $n = 3$ biological replicates of pooled FAPs per condition (**b**, **d**, **f–i**). In **b** and **d**, all conditions were processed together, so that the control in **b** is identical to the control in **d**. Error bars represent s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; unpaired Student's t -test (**b**, **d**) or paired ratio t -test (**f–i**).

Fig. 3c). Conversely, pre-treatment of cells with the pA-AMOs resulted in increased ERK1/2 phosphorylation (Fig. 2h, i).

These changes in signalling were also accompanied by changes in FAP activation. FAPs treated with the pA-AMOs exhibited an increased proliferative response to PDGF-AA (Fig. 3a and Extended Data Fig. 4a–c), as did cells treated with an In-PDGFR α -specific short interfering RNA (siRNA) (Extended Data Figs 4d, 5a–d). By contrast, upregulation of In-PDGFR α by 5'ss-AMO pre-treatment blunted PDGF-induced proliferation (Fig. 3b and Extended Data Fig. 4e), as did the overexpression of In-PDGFR α (Extended Data Fig. 4f). In independent functional assays, we found that pre-treatment

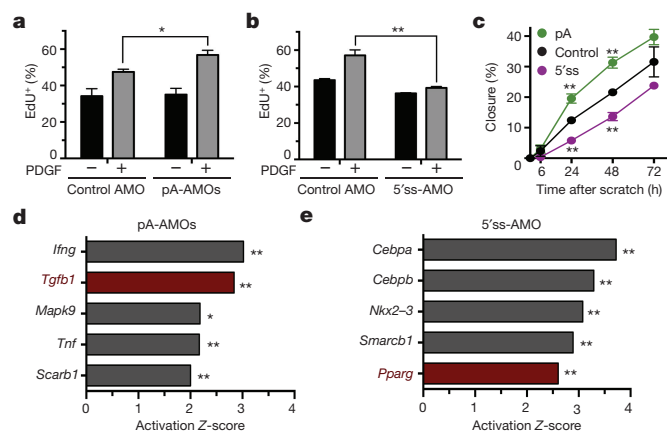


Figure 3 | Intronic polyadenylation of *Pdgfra* limits activation of FAPs *in vitro*, and reducing levels of the intronic variant promotes FAP activation. **a, b**, Treatment of FAPs with pA-AMOs increases FAP proliferation in response to PDGF-AA (**a**) whereas 5'ss-AMO treatment blunts proliferation (**b**). **c**, FAPs treated with pA-AMOs show enhanced proliferation and migration whereas those treated with the 5'ss-AMO displayed delayed proliferation and migration in scratch assays. **d, e**, Ingenuity pathway analysis shows that *Tgfb1*, a promoter of fibrosis, is a top predicted regulator of gene expression change in FAPs treated with pA-AMOs compared to control treatment (**d**) whereas treatment with the 5'ss-AMOs activates genes implicated in reducing fibrosis, including *Pparg* (**e**). **a–c**, $n = 3$ biological replicates of pooled FAPs per condition or time point. Error bars represent s.e.m. * $P < 0.05$, ** $P < 0.01$; unpaired Student's *t*-tests (**a–c**). In **d** and **e**, two pooled FAP samples per condition were used with the overlap *P* value calculated using the Fisher's exact test across genes and gene sets.

with pA-AMOs increased FAP proliferation and migration whereas 5'ss-AMO pre-treatment resulted in a decrease (Fig. 3c).

We also investigated the effects of In-PDGFR α modulation on FAP differentiation. FAPs treated with PDGF-AA showed upregulation of key fibrosis markers and induction of downstream TGF β signalling (Extended Data Fig. 6a, b), consistent with previous studies¹². Notably, gene expression of FAPs treated with pA-AMOs, in which In-PDGFR α levels are decreased, revealed a pattern consistent with enhanced activation and fibrotic differentiation. In particular, there was enrichment for DNA replication and cell cycle genes including the MAPK-ERK pathway (Extended Data Fig. 6c–e). Moreover, causal network analysis suggested that TGF β 1 signalling was active in these cells (Fig. 3d) with increased expression of associated fibrosis mediators including connective tissue growth factor. Conversely, the inhibition of PDGFR α signalling through 5'ss-AMO-mediated In-PDGFR α upregulation did not significantly change gene expression associated with TGF β 1 activation. Instead, there was enrichment for processes related to protein and/or RNA processing and metabolism (Extended Data Fig. 6f–h). The top predicted regulators included *Pparg* (Fig. 3e); the activity of this gene is associated with a reduction in fibrosis and TGF β signalling in several tissues²².

Given these findings and the association of FAPs with fibrosis, we hypothesized that altering PDGF signalling in FAPs by modulating In-PDGFR α levels *in vivo* would affect muscle fibrosis following injury. Therefore, we designed PDGFR α -specific vivo-morpholinos (VMOs), a class of morpholinos that can enter cells directly because of a covalently bound delivery moiety²³. The VMOs targeting the polyadenylation site (pA-VMOs) and the 5'-splice-site (5'ss-VMO) were designed to contain the same targeting sequences as their AMO counterparts. We tested these VMOs both *in vitro* and *in vivo* and observed that their effects on In-PDGFR α and FL-PDGFR α levels mimicked those of their corresponding AMOs (Fig. 4a, b and Extended Data Fig. 7a, b). Using intramuscular injection of VMOs in mice, we found that the 5'ss-VMO resulted in a decreased FAP proliferation index and a corresponding decrease in FAP numbers (Extended

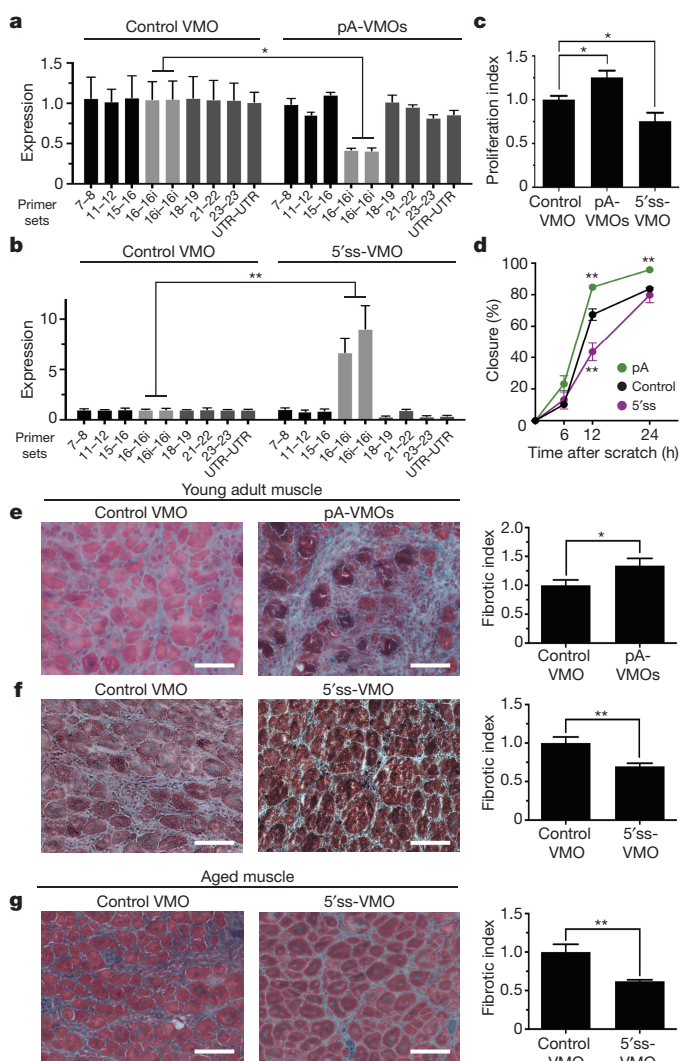


Figure 4 | Downregulation of In-PDGFR α enhances FAP activation and fibrosis, whereas enhancing intronic polyadenylation of PDGFR α attenuates FAP activation and fibrosis. **a, b**, FAPs treated *in vivo* with pA-VMOs show a downregulation of the In-PDGFR α transcript ($n = 3$) (**a**), whereas those treated with the 5'ss-VMO exhibit an upregulation of In-PDGFR α (control, $n = 4$; 5'ss-VMO, $n = 3$) (**b**). **c, d**, FAPs treated *in vivo* with pA-VMOs show enhanced proliferation and migration *ex vivo* as measured by EdU incorporation (**c**) and a scratch assay (**d**), whereas those treated with the 5'ss-VMO show a decrease in proliferation (**c**; control, $n = 6$; pA-VMO, $n = 5$; 5'ss-VMO, $n = 5$) and migration (**d**; $n = 5$ for all treatments). **e, f**, Representative images of Gomori-trichrome stained cryosections and quantification of fibrosis of glycerol-injured muscles treated with pA-VMOs (control, $n = 9$; pA-VMO, $n = 10$) (**e**) or 5'ss-VMO (control, $n = 9$; 5'ss-VMO, $n = 10$) (**f**) compared to the control VMO in young adult mice. pA-VMOs and the 5'ss-VMO show increased and decreased fibrosis levels compared to the control VMO, respectively. **g**, Reduced fibrosis in barium chloride-injured muscle by treatment with the 5'ss-VMO in aged mice (control, $n = 9$; 5'ss-VMO, $n = 10$). The fibrotic index is defined as the area of fibrosis divided by total area of muscle, normalized to control-treated muscle. n represents biological replicates of pooled FAPs (**a–d**) or individual tibialis anterior muscles (**e, f**). Scale bars, 100 μ m. Error bars represent s.e.m. * $P < 0.05$, ** $P < 0.01$; unpaired Student's *t*-tests.

Data Fig. 7c, d). Treatment with pA-VMOs did not enhance FAP proliferation significantly, possibly because cells are already maximally stimulated by the injury response *in vivo* (Extended Data Fig. 7c, d).

In parallel to our *in vitro* analyses, FAPs isolated from muscle pre-treated with pA-VMOs displayed increased proliferation and enhanced migration (Fig. 4c, d). Moreover, expression analysis of these

cells revealed enrichment for terms related to proliferation including cell cycle and mitosis (Extended Data Fig. 7e). Notably, causal network analysis was significant ($P < 0.05$) for several proliferation and fibrosis regulators including connective tissue growth factor, a growth factor that is upregulated by TGF β and has been shown to increase collagen production by fibroblasts and to promote tissue fibrosis⁸ (Extended Data Fig. 7g). By contrast, 5'ss-VMO-treated FAPs showed decreased proliferation and migration (Fig. 4c, d) as well as enrichment for interferon signalling (Extended Data Fig. 7f, h), which has been shown to decrease fibroblast activation and reduce fibrosis levels in models of hepatic and renal injury²⁴.

We hypothesized that enhancing FAP activity through inhibition of In-PDGFR α via intramuscular injection of pA-VMOs would enhance fibrosis following injury. Indeed, this occurred in two separate models of muscle injury (Fig. 4e and Extended Data Fig. 8). Although FAPs have been implicated as key mediators of fibrosis, the ability of pA-VMOs to promote fibrosis, albeit with a more modest effect on FAP proliferation *in vivo* (Extended Data Fig. 7c, d), could reflect the contribution of other cell types to the fibrotic response. We next investigated whether the 5'ss-VMO, because of its potential to block PDGFR α signalling, could decrease fibrosis. Notably, not only did injection of the 5'ss-VMO decrease fibrosis following injury in young adult muscle (Fig. 4f and Extended Data Fig. 8), but also overall levels of fibrosis were significantly reduced ($P < 0.01$) in aged muscle (Fig. 4g), where this type of pathology is more pronounced⁵.

Our finding that *Pdgfra* is subject to intronic polyadenylation and that this process influences FAP activation suggests that polyadenylation site choice can serve as a mechanism by which stem cells adjust levels of molecular signalling. VMOs are attractive candidates for disease treatment because of their sequence specificity, stability, resistance to degradation and lack of interaction with cellular structures²⁵. Their utility in reducing tissue fibrosis through experimental manipulation of *Pdgfra* intronic variant levels suggests that modulating polyadenylation site selection could be a promising therapeutic approach to the treatment of fibrotic diseases and disorders.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 July 2014; accepted 17 October 2016.

Published online 28 November 2016.

- Orr-Urtreger, A., Bedford, M. T., Do, M. S., Eisenbach, L. & Lonai, P. Developmental expression of the α receptor for platelet-derived growth factor, which is deleted in the embryonic lethal Patch mutation. *Development* **115**, 289–303 (1992).
- Soriano, P. The PDGFR α receptor is required for neural crest cell development and for normal patterning of the somites. *Development* **124**, 2691–2700 (1997).
- Andrae, J., Gallini, R. & Betsholtz, C. Role of platelet-derived growth factors in physiology and medicine. *Genes Dev.* **22**, 1276–1312 (2008).
- Olson, L. E. & Soriano, P. Increased PDGFR α activation disrupts connective tissue development and drives systemic fibrosis. *Dev. Cell* **16**, 303–313 (2009).
- Brack, A. S. *et al.* Increased Wnt signaling during aging alters muscle stem cell fate and increases fibrosis. *Science* **317**, 807–810 (2007).
- Goldspink, G., Fernandes, K., Williams, P. E. & Wells, D. J. Age-related changes in collagen gene expression in the muscles of mdx dystrophic and normal mice. *Neuromuscul. Disord.* **4**, 183–191 (1994).
- Conboy, I. M. & Rando, T. A. Aging, stem cells and tissue regeneration: lessons from muscle. *Cell Cycle* **4**, 407–410 (2005).
- Mann, C. J. *et al.* Aberrant repair and fibrosis development in skeletal muscle. *Skelet. Muscle* **1**, 21 (2011).
- Serrano, A. L. *et al.* Cellular and molecular mechanisms regulating fibrosis in skeletal muscle repair and disease. *Curr. Top. Dev. Biol.* **96**, 167–201 (2011).

- Uezumi, A., Fukada, S., Yamamoto, N., Takeda, S. & Tsuchida, K. Mesenchymal progenitors distinct from satellite cells contribute to ectopic fat cell formation in skeletal muscle. *Nat. Cell Biol.* **12**, 143–152 (2010).
- Joe, A. W. B. *et al.* Muscle injury activates resident fibro/adipogenic progenitors that facilitate myogenesis. *Nat. Cell Biol.* **12**, 153–163 (2010).
- Uezumi, A. *et al.* Fibrosis and adipogenesis originate from a common mesenchymal progenitor in skeletal muscle. *J. Cell Sci.* **124**, 3654–3664 (2011).
- Fiore, D. *et al.* Pharmacological blockage of fibro/adipogenic progenitor expansion and suppression of regenerative fibrogenesis is associated with impaired skeletal muscle regeneration. *Stem Cell Res.* **17**, 161–169 (2016).
- Tidball, J. G., Spencer, M. J. & St Pierre, B. A. PDGF-receptor concentration is elevated in regenerative muscle fibers in dystrophin-deficient muscle. *Exp. Cell Res.* **203**, 141–149 (1992).
- Zhao, Y. *et al.* Platelet-derived growth factor and its receptors are related to the progression of human muscular dystrophy: an immunohistochemical study. *J. Pathol.* **201**, 149–159 (2003).
- Ito, T. *et al.* Imatinib attenuates severe mouse dystrophy and inhibits proliferation and fibrosis-marker expression in muscle mesenchymal progenitors. *Neuromuscul. Disord.* **23**, 349–356 (2013).
- Huang, P., Zhao, X. S., Fields, M., Ransohoff, R. M. & Zhou, L. Imatinib attenuates skeletal muscle dystrophy in mdx mice. *FASEB J.* **23**, 2539–2548 (2009).
- Ozsolak, F. *et al.* Direct RNA sequencing. *Nature* **461**, 814–818 (2009).
- Boutet, S. C. *et al.* Alternative polyadenylation mediates microRNA regulation of muscle stem cell function. *Cell Stem Cell* **10**, 327–336 (2012).
- Mueller, A. A., Cheung, T. H. & Rando, T. A. All's well that ends well: alternative polyadenylation and its implications for stem cell biology. *Curr. Opin. Cell Biol.* **25**, 222–232 (2013).
- Vorlová, S. *et al.* Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. *Mol. Cell* **43**, 927–939 (2011).
- Kawai, T. *et al.* PPAR γ agonist attenuates renal interstitial fibrosis and inflammation through reduction of TGF β . *Lab. Invest.* **89**, 47–58 (2009).
- Morcos, P. A., Li, Y. & Jiang, S. Vivo-Morpholinos: a non-peptide transporter delivers Morpholinos into a wide array of mouse tissues. *Biotechniques* **45**, 613–614, 616, 618 passim (2008).
- Bansal, R. *et al.* Novel engineered targeted interferon γ blocks hepatic fibrogenesis in mice. *Hepatology* **54**, 586–596 (2011).
- Du, L. & Gatti, R. A. Potential therapeutic applications of antisense morpholino oligonucleotides in modulation of splicing in primary immunodeficiency diseases. *J. Immunol. Methods* **365**, 1–7 (2011).
- Heinrich, M. C. *et al.* PDGFRA activating mutations in gastrointestinal stromal tumors. *Science* **299**, 708–710 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank members of the Rando laboratory, especially A. de Morree, M. Quarta, J. Shih, B. Yoo, V. Garcia and I. Akimenko for discussions and experimental assistance as well as L. Rott for guidance with FACS. This work was supported by the Glenn Foundation for Medical Research, by grants from the National Institutes of Health (F30 AG043235) and the California Institute for Regenerative Medicine (TG2-01159) to A.A.M., and by grants from the Department of Veterans Affairs and the NIH (P01 AG036695, R01 AG23806, R01 AR062185, and TR01 AG47820) to T.A.R.

Author Contributions A.A.M. and T.A.R. conceived the study and were involved in the overall design of experiments. A.A.M. and T.H.C. designed, conducted and analysed the microarray and sequencing experiments as well as the FACS analyses. A.A.M. designed the morpholinos and siRNAs and also performed and analysed the associated experiments to test for their transcriptional effects. A.A.M. and C.T.V.V. designed the viral construct while C.T.V.V. created the construct and performed experiments to analyse expression. A.A.M. and C.T.V.V. designed and performed experiments and data analysis for PDGFR α protein expression and downstream signalling in response to morpholino treatment *in vitro* and *in vivo* as well as for muscle fibrosis responses. A.A.M. conducted and analysed the proliferation assays whereas C.T.V.V. performed and evaluated the scratch assays. K.D.F. carried out pilot experiments to assess for PDGFR α signalling in regenerating skeletal muscle and in initial FACS characterization of the FAP population. A.A.M. and T.A.R. wrote the manuscript with input from C.T.V.V. and T.H.C.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.A.R. (rando@stanford.edu).

Reviewer Information Nature thanks F. Rossi, Y. Shi and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Animals. Wild-type male C57BL/6 mice and B6.129S4-PDGFR α ^{tm1.1(EGFP)Sor/J} mice (Jackson strain number 007669), which contain an H2B-eGFP fusion protein knocked into the *Pdgfra* locus, were obtained from Jackson Laboratories. Young adult mice were 6–8 weeks of age; aged mice were 22–24 months of age. Mice were housed and maintained in the Veterinary Medical Unit at the Veterans Affairs Palo Alto Health Care System. Animal protocols were performed in accordance with the policies of the Administrative Panel on Laboratory Animal Care of Stanford University.

Muscle injury. Mice were anaesthetized using isoflurane. To assess muscle regeneration, 50 μ l of a 1.2% barium chloride (BaCl₂) solution (Sigma-Aldrich) was injected into tibialis anterior muscles as described previously⁵. To isolate activated FAPs for western blot analysis and FACS analysis, 50 μ l of 1.2% BaCl₂ or 50% (v/v) glycerol/water was injected throughout the lower hindlimb muscles. For induction of fibrosis, 30 μ l of 50% (v/v) glycerol or 30 μ l 1.2% BaCl₂ solution was injected into tibialis anterior muscles.

Isolation of FAPs. Muscles were dissected from mice and dissociated mechanically. All hindlimb muscles were used except in experiments where FAPs were isolated from VMOs injected into tibialis anterior muscles. In this case, only the tibialis anterior muscle was dissected. The muscle suspension was digested using collagenase II (760 U ml⁻¹; Worthington Biochemical Corporation) in Ham's F10 medium (Invitrogen) with 10% horse serum (Invitrogen) for 90 min at 37°C with agitation. The suspension was then washed and digested in collagenase II (152 U ml⁻¹; Worthington Biochemical Corporation) and dispase (2 U ml⁻¹; Invitrogen) for 30 min at 37°C with agitation. The resultant mononuclear cells were then stained with the following antibodies: VCAM-1-biotin (clone 429; BioLegend, 105704), CD31-APC (clone MEC 13.3; BioLegend, 102510), CD45-APC (clone 30-F11; BioLegend, 103112) and Sca-1-Pacific Blue (clone D7; BioLegend, 108120) at 1:75. Streptavidin-PE-Cy7 (BioLegend, 405206) at 1:75 was used to amplify the VCAM-1 signal. FAPs were collected according to the following sorting criteria: CD31⁺CD45⁺Sca-1⁺. FACS was performed using BD-FACS Aria II and BD-FACS Aria III cell sorters equipped with 488 nm, 633 nm and 405 nm lasers. The cell sorters were carefully optimized for purity and viability and sorted cells were subjected to FACS analysis immediately after sorting to confirm FAP purity.

Direct RNA sequencing. FAPs were isolated from uninjured C57BL/6 mice as described above and lysed. RNA was prepared with the RNeasy Mini Kit as per the manufacturer's instructions (Qiagen). A 3' blocking reaction was performed using a poly(A) tailing kit (Ambion) and 3'-dATP (Jena Bioscience) and the reaction mixture was incubated at 37°C for 30 min. RNA was hybridized to flow cell surfaces for direct RNA sequencing as previously described¹⁸. Raw direct RNA sequencing reads were filtered using the Helicos-developed pipeline, Helisphere, to eliminate reads less than 25 nucleotides long or of low quality. These reads were then mapped to the mouse genome (NCBI37/mm9) using an IndexDPgenomic module and reads with a score above 4.3 were allowed. To avoid artefacts from mispriming, reads mapping to regions in the genome where more than four consecutive adenines were coded immediately 3' to the mapping sequence were excluded from further analysis. Reads were viewed using the Integrative Genomics Viewer^{32,33}.

Analysis of alternative polyadenylation sites and quantitative RT-PCR. Total RNA was extracted from FAPs isolated from uninjured C57BL/6 mice using TRIzol (Invitrogen) as per the manufacturer's instructions. To identify the polyadenylation sites, the sample was reverse transcribed using the SMARTer RACE cDNA amplification kit (Clontech) according to the manufacturer's instructions using the primers listed in Extended Data Table 1. The amplified fragments were subcloned into pGEM-T-Easy (Promega) and sequenced. Sequencing data were visualized with 4Peaks. To assess levels of the intronic variant and UTR variants, primers were designed to span the *Pdgfra* transcript (Extended Data Table 2). Variant expression was normalized to *Gapdh* using the comparative C_t method²⁷ and reported relative to the average of control-treated samples.

Retroviral transfection. A construct corresponding to In-PDGFR α (DNAFORM, AK035501, RIKEN clone 9530057A20) was obtained. This construct was subcloned into the pMXs-IRES-GFP retroviral backbone (Cell BioLabs, Inc.) to generate pMXs-I-P α . Replication-incompetent retroviral particles were generated by transfection of the 293T human embryonic kidney cell-derived Phoenix helper cell line (gift from G. Nolan). Viral supernatant was filtered through 0.45- μ m polyethersulfone filters, concentrated using PEG precipitation and stored at -80°C.

FAPs were plated in 6-well plates and grown in DMEM supplemented with 10% fetal bovine serum (FBS). When cells reached 70% confluency, viral supernatant and polybrene (at a final concentration of 4 μ g ml⁻¹) were added to the medium. For overexpression experiments, FAPs were incubated with the viral supernatant for 48 h before analysis. For signalling assays, FAPs were incubated with the viral supernatant for 24 h. Afterwards the medium was changed to serum-free DMEM containing viral supernatant and the cells were incubated for an additional 24 h. The FAPs were then treated with 1 ng ml⁻¹ PDGF-AA for 15 min, after which the cells were used for western blot analysis.

Generation of an In-PDGFR α -specific rabbit polyclonal antibody. A peptide with the sequence GKSAHAHSGKYDLSVV, which represents the unique C-terminal region of In-PDGFR α protein, was generated (Thermo Scientific Pierce, OE0726). To generate In-PDGFR α rabbit polyclonal antibodies directed against In-PDGFR α , New Zealand white rabbits that were specific pathogen free were immunized with 0.25 mg of the peptide in Complete Freund's Adjuvant. The rabbits received three boosters of antigen consisting of 0.10 mg in Incomplete Freund's Adjuvant at days 14, 42 and 56 after immunization. Serum was collected at days 70 and 72 (Thermo Scientific Pierce).

Western blot analysis. Cells and homogenized tissues were lysed with RIPA lysis buffer supplemented with protease and phosphatase inhibitors (Roche). The lysates were run on Criterion SDS-PAGE gels (Bio-Rad), transferred to nitrocellulose membranes (Fisher Scientific), and analysed by western blot using the following rabbit antibodies: PDGFR α polyclonal (1:1,000, Cell Signaling, 3174), PDGFR α centre (1:100, Abgent, AP14254c), In-PDGFR α custom (1:1,000), pPDGFR α ^{Tyr754} polyclonal (1:1,000, Cell Signaling, 4547), Akt polyclonal (1:1,000, Cell Signaling, 9272), pAkt polyclonal (1:1,000, Cell Signaling, 9271), PLC γ polyclonal (1:1,000, Cell Signaling, 5690), pPLC γ polyclonal (1:1,000, Cell Signaling, 2821), ERK polyclonal (1:2,000, Cell Signaling, 4695), pERK polyclonal (1:2,000, Cell Signaling, 4370), SMAD2/3 monoclonal (1:1,000, Cell Signaling, 8685), and pSMAD2^{Ser465/Ser467}/SMAD3^{Ser423/Ser425} monoclonal (1:1,000, Cell Signaling, 8828). Membranes were incubated in horseradish-peroxidase-labelled secondary antibodies and bands were visualized with enhanced chemiluminescence (Advansta).

siRNA knockdown of PDGFR α variants. siRNAs were designed using the Dharmacon siDESIGN Center for knockdown of In-PDGFR α and FL-PDGFR α (Extended Data Table 2). To knockdown either In-PDGFR α or FL-PDGFR α in FAPs, approximately 8×10^4 cells were plated in a 12-well plate containing DMEM supplemented with 10% FBS and grown to 70–80% confluence. Cells were incubated in 200 nM of either PDGFR α or control siRNAs using Lipofectamine 2000 (Invitrogen). To assess knockdown, cells were collected at 24 h for qPCR analysis. For western blot analyses, 3×10^5 cells were plated in 6-well plates and incubated in Ham's F10 medium (Invitrogen) supplemented with 10% horse serum (Invitrogen) for 24 h. The medium was then replaced with serum-free Ham's F10 (Invitrogen) supplemented with 200 nM siRNA and incubated for an additional 24 h.

Antisense morpholino treatment. Morpholinos were designed to target two polyadenylation sites on the intronic variant (pA₁: 5'-TGATTACAT TATATCTGTCTTTATT-3' and pA₂: 5'-AGCAAAGACCATCATAGCAGA ATGA-3') and the upstream 5' splice site of the intron (5'ss: 5'-ATGGGCACTT TTACCTAGCATGGAT-3') (Gene Tools, LLC). For *in vitro* treatment, cells were grown to 70–80% confluency in DMEM (Invitrogen) supplemented with 10% FBS (Atlanta Biologicals). Cells were incubated in 10 μ M of the indicated morpholino using the Endo-Porter transfection reagent (Gene Tools, LLC). Cells were collected at 24 h for qPCR analysis with RNA isolated using the RNeasy Plus Mini kit with on-column DNase digestion as per manufacturer's instructions (Qiagen). For western blot analysis, cells were transfected for 24 h in Ham's F10 medium (Invitrogen) supplemented with 10% horse serum (Invitrogen). The medium was then replaced with serum-free Ham's F10 (Invitrogen) and incubated for an additional 24 h. For signalling assays, cells were then incubated for 15 min with PDGF-AA (Peprotech) at 0.1 ng ml⁻¹ or 20 ng ml⁻¹ for cells treated with pA-AMOs or 5'ss-AMO, respectively, and lysed for western blot analysis as described above.

In vitro EdU incorporation assay. For AMO treatment, FAPs were isolated from the uninjured hindlimb muscles of C57BL/6 mice and seeded at 1×10^5 cells per well in poly-D-lysine-coated 8-well chamber slides (BD Biosciences) coated with ECM gel (Sigma-Aldrich). Cells were transfected with 10 μ M AMO using Endoport (Gene Tools) and expanded for 2 days in Ham's F10 (Invitrogen) supplemented with 10% horse serum (Invitrogen). The medium was then replaced with Opti-MEM supplemented with 2 ng ml⁻¹ PDGF-AA ligand and 10 μ M EdU (Invitrogen). Cells were fixed in 4% paraformaldehyde (Sigma-Aldrich) after 24 h.

For siRNA treatment, FAPs were isolated from the uninjured hindlimb muscles of C57BL/6 mice and seeded at 2×10^5 cells per well in poly-D-lysine coated 8-well chamber slides (BD Biosciences) coated with ECM gel (Sigma-Aldrich). The medium was supplemented with 200 nM siRNA and transfected using Lipofectamine 2000 (Invitrogen). After 24 h, the medium was replaced with Opti-Mem and the cells were re-transfected with 200 nM siRNA and 50 ng ml⁻¹ PDGF-AA. In siRNA-treated samples, EdU was not included in this medium. Rather, after 20 h the medium was replaced with Opti-Mem containing 10 μ M EdU (Invitrogen). Cells were fixed 4 h later.

For retroviral overexpression of In-PDGFR α , FAPs were isolated from uninjured hindlimbs of C57BL/6 mice and seeded at 2×10^5 cells per well in poly-D-lysine coated 8-well chamber slides (BD Biosciences) coated with ECM gel (Sigma-Aldrich). FAPs were cultured in DMEM supplemented with 10% FBS along with viral supernatant and 4 μ g ml⁻¹ polybrene. After 24 h, the medium was

replaced with serum-free DMEM containing viral supernatant and 20 ng ml⁻¹ PDGF-AA. Twenty hours later, the medium was replaced with Opti-MEM containing 10 μ M EdU. Cells were fixed after 4 h. For EdU incorporation experiments, cells were stained using the Click-iT EdU Imaging Kit (Invitrogen). Cells were analysed on a Zeiss Observer Z1 fluorescent microscope (Carl Zeiss) equipped with a Hamamatsu Orca-ER camera (Hamamatsu) and Improvision Volocity software (Perkin Elmer).

Scratch assay. Cells isolated by FACS from uninjured hindlimb muscles were seeded at a density of 3.5×10^4 cells per well in 96-well plates in Ham's F10 medium supplemented with 2% horse serum. After 48 h, cells were nearly confluent and the medium was changed to Ham's F10 with 2% horse serum and 20 ng ml⁻¹ PDGF-AA. A wound was made by scratching a 200- μ l pipette tip across the monolayer of cells. The initial scratch area was determined immediately and set to 100%. Images were taken at regular intervals and the scratch area at each time point was measured and calculated as a percentage of the initial scratch area. Scratch closure is defined as the inverse of the cell-free area as a percentage of total area.

Microarray analysis. For *in vitro* microarray analysis, FAPs were isolated from the uninjured hindlimb muscles of C57BL/6 mice. Cells were plated at 1×10^6 cells per well in 12-well plates. Cells were grown for 2.5 days in DMEM supplemented with 10% FBS. The medium was switched to Ham's F10 supplemented with 10% horse serum and transfected with 10 μ M AMO as indicated for 48 h. The medium was then replaced with Opti-Mem and cells were re-transfected with 10 μ M AMO. After 48 h, the cells were lysed and RNA was prepared with the RNeasy Mini Kit as per the manufacturer's instructions (Qiagen). For *in vivo* microarray analysis, tibialis anterior muscles were injured with 30 μ l of glycerol each and injected with the indicated VMO after 3 days. FAPs were then isolated from the muscles 2 days after VMO injection. Cells were pelleted and RNA prepared from samples as indicated above. The microarray data were obtained using Affymetrix Mouse 1.0 ST. For gene set enrichment analysis (GSEA), the samples were normalized and processed using GenePattern ExpressionFileCreator and PreProcessData set modules. Expression data were analysed and visualized with GSEA²⁸ and GENE-E (<http://www.broadinstitute.org/cancer/software/GENE-E/>). For ingenuity pathway analysis, including causal network analysis, the samples were normalized using Affymetrix Expression Console Software and analysed for enrichment using IPA (Ingenuity Systems, <http://www.ingenuity.com>). Array data were deposited into Gene Expression Omnibus (Accessions GSE60099 and GSE81744).

Vivo-morpholino treatment. Vivo-morpholinos were designed to target two polyadenylation sites on the intronic variant (pA₁-VMO: 5'-TGATTACATTAT ATCTGTCTTTATT-3' and pA₂-VMO: 5'-AGCAAAGACCATCATAGC AGAATGA-3') and the upstream 5' splice site of the intron (5'ss-VMO: 5'-ATG GGCACCTTTACCTAGCATGGAT-3') (Gene Tools, LLC). For treatment *in vitro*, cells were isolated from hindlimb muscles of C57BL/6 mice and grown to 70–80% confluency in DMEM (Invitrogen) supplemented with 10% FBS (Atlanta Biologicals). Cells were incubated in the 10 μ M of the indicated morpholino (Gene Tools, LLC). Cells were collected at 24 h for qPCR analysis. For *in vivo* qPCR analysis, tibialis anterior muscles were injured with glycerol as described above and injected with 250 ng of the indicated VMO at the site of injury 3 days later. FAPs were sorted by FACS 7 days after VMO injection for qPCR analysis.

For *ex vivo* proliferation and scratch assays, tibialis anterior muscles were injured with glycerol and injected with 250 ng of the indicated VMO 3 days after injury. FAPs were isolated 2 days later by FACS. In EdU incorporation studies, cells were

seeded at 4×10^4 cells per well in poly-D-lysine-coated 8-well chamber slides (BD Biosciences) coated with ECM gel (Sigma-Aldrich). Cells were incubated in 10 ng ml⁻¹ PDGF-AA (Peprotech) and 10 μ M EdU (Invitrogen) for 24 h. The cells were fixed and stained. In the *ex vivo* proliferation studies as well as the *in vivo* proliferation studies described below, the proliferation index was used to denote the percentage EdU incorporation normalized to control. In the scratch assays, cells were seeded and treated as described above.

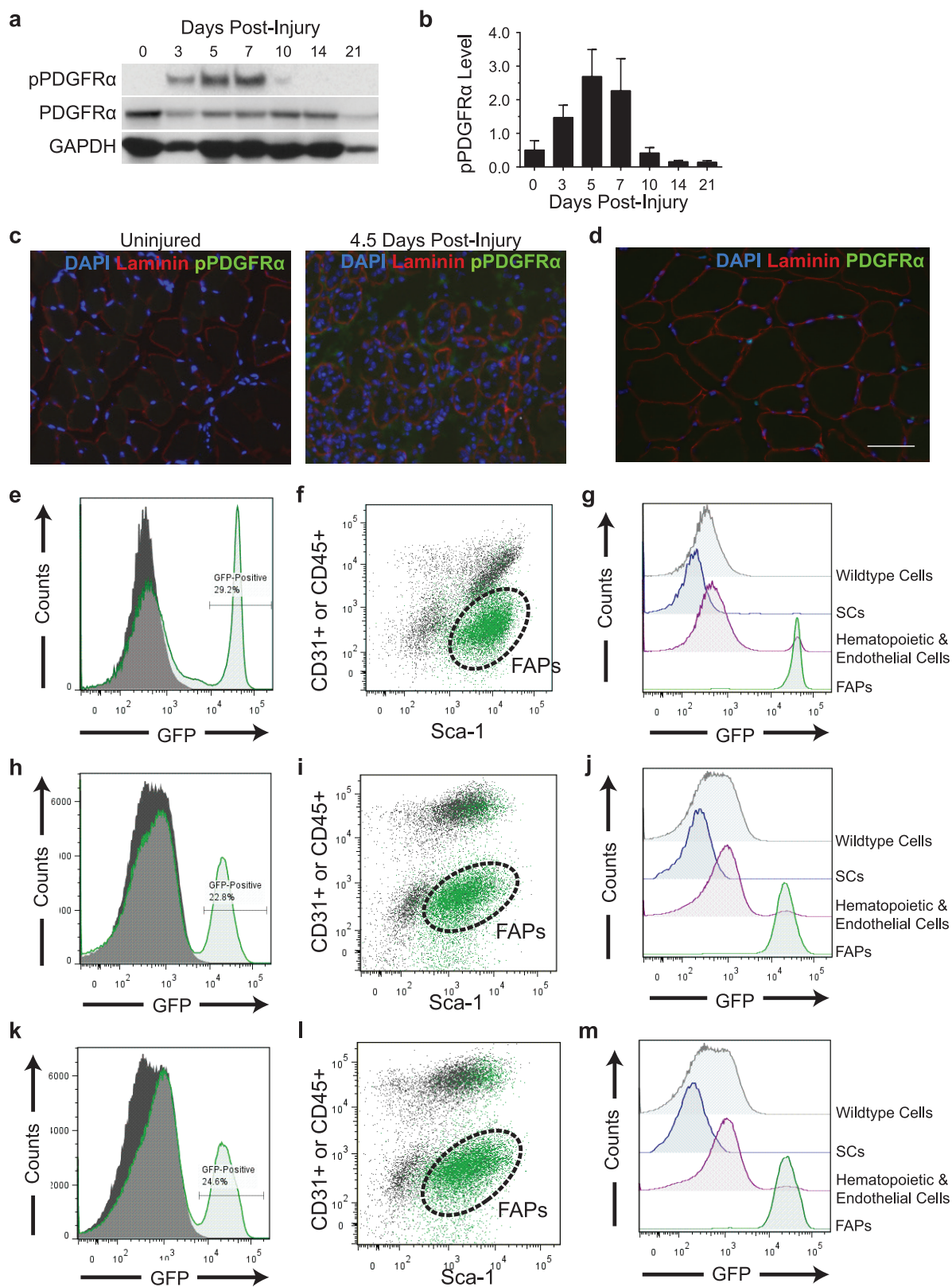
For *in vivo* proliferation studies, tibialis anterior muscles were injected with 150 ng of the indicated VMO at 0 and 24 h. FAPs were isolated at 48 h via FACS. To assess *in vivo* proliferation, the cells were exposed to 10 μ M EdU immediately after muscle isolation and incubated in 10 μ M EdU *ex vivo* during the collagenase, collagenase/dispase, and antibody incubations as described above. The cells were plated in poly-D-lysine-coated 8-well chamber slides (BD Biosciences) coated with ECM gel (Sigma-Aldrich), fixed 1 h after plating, and stained using the Click-iT EdU Imaging Kit (Invitrogen).

For histological analysis, tibialis anterior muscles were injured with glycerol or BaCl₂ and injected at the site of injury with 250 ng of the indicated VMO. After 7 days, the muscles were snap frozen in isopentane cooled in liquid nitrogen immediately after dissection. Muscles sections were stained with Gomori-trichrome (Richard-Allan Scientific) per manufacturer's instructions or oil red O (Sigma-Aldrich) as previously described²⁹. The fibrotic index was calculated as the area of fibrosis divided by total area of muscle normalized to control-treated muscle. The fibro-adipose index was defined as the area of fibrosis plus the area of adiposis (as detected by oil red O staining) divided by total area of muscle, normalized to control.

Statistics and general methods. Major factors in determining sample size included the level of the effect and the inherent variability in measurements obtained. No statistical methods were used to predetermine sample size. Animals were excluded from the study only if their health status was compromised, such as occurred when animals had visible wounds from fighting. Samples were not specifically randomized or blinded. However, mouse identifiers were used when possible to blind evaluators to experimental conditions, and all samples within experiments were processed identically for measurement quantification using automated tools as specified.

Data availability. The sequencing data were deposited into the NCBI Sequence Read Archive (accession number SRP079186). Array data were deposited into Gene Expression Omnibus (accession numbers GSE60099 and GSE81744).

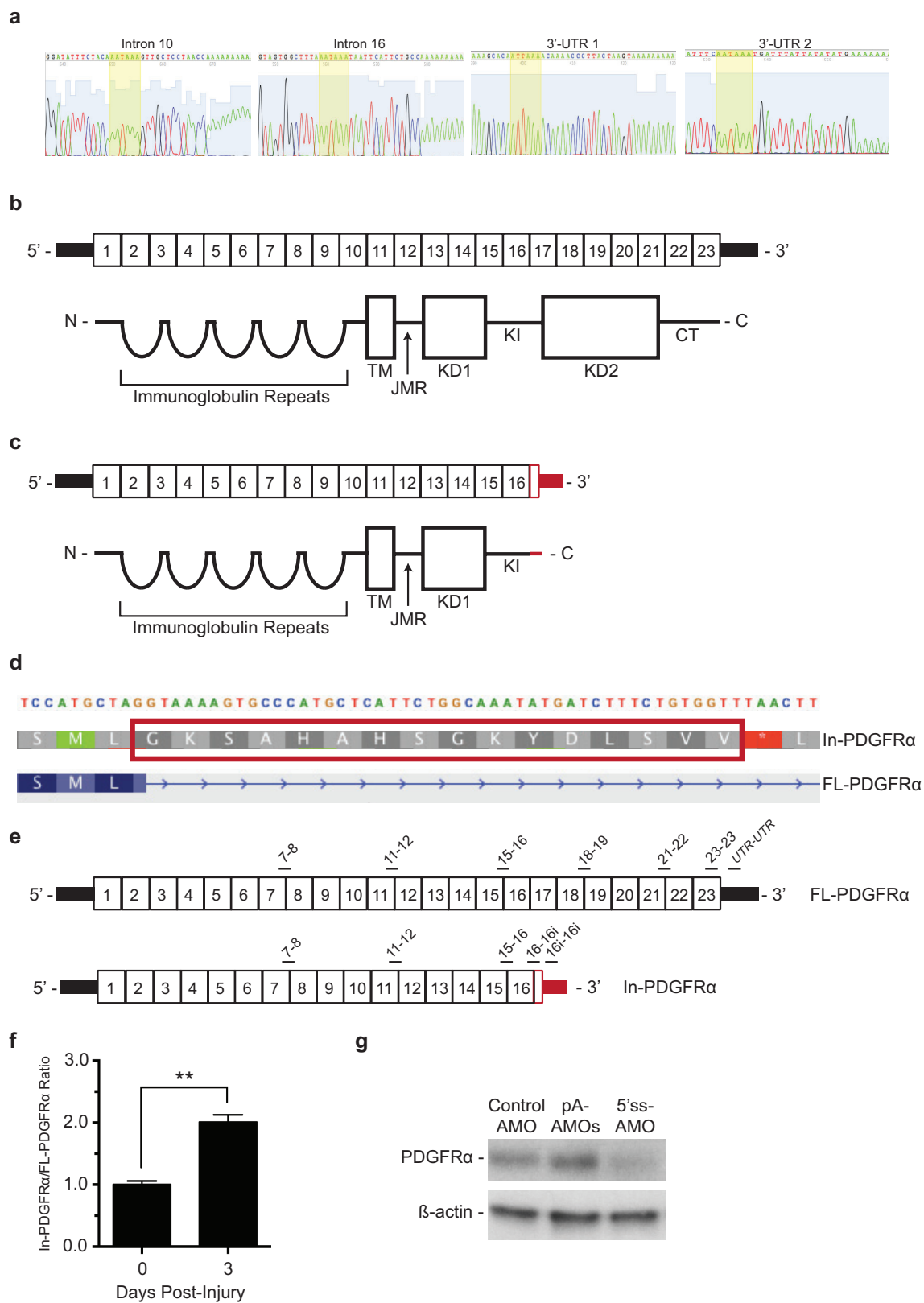
27. Pfaffl, M. W., A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).
28. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
29. Mehlem, A., Hagberg, C. E., Muhl, L., Eriksson, U. & Falkevall, A. Imaging of neutral lipids by oil red O for analyzing the metabolic status in health and disease. *Nat. Protoc.* **8**, 1149–1154 (2013).
30. Milacic, M. *et al.* Annotating cancer variants and anti-cancer therapeutics in reactome. *Cancers* **4**, 1180–1211 (2012).
31. Croft, D. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–D477 (2014).
32. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
33. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).



Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | PDGFR α is activated specifically in FAPs during muscle regeneration. **a**, Tibialis anterior muscles were dissected on the indicated days after injury. The muscle lysates were subjected to western blot to probe for PDGFR α and pPDGFR α using different antibodies, which shows increased expression of pPDGFR α during muscle regeneration. **b**, Quantification of **a** ($n = 3$ individual tibialis anterior muscles per time point). pPDGFR α level refers to expression levels for pPDGFR α normalized to total PDGFR α . Error bars represent s.e.m. **c**, Immunofluorescence of an uninjured tibialis anterior muscle and a muscle 4.5 days after injury shows an increase in the abundance of pPDGFR α -expressing cells during the regenerative process. (DAPI, blue; laminin, red; pPDGFR α , green). **d**, Immunofluorescence of an uninjured tibialis anterior muscle from a PDGFR α -eGFP reporter mouse. **e**, FACS

histogram analysis of the GFP signal in a PDGFR α -eGFP reporter heterozygote (green line) and in a wild-type littermate (grey, solid) in cells from uninjured hindlimb muscles. **f**, FACS plot of cells isolated from uninjured hindlimb muscles of the PDGFR α -eGFP reporter. GFP-positive cells (shown in green) are overlaid upon all cells (shown in grey). Dashed lines represent the gate for the FAP population. **g**, Histograms showing GFP signal in cell populations isolated from uninjured PDGFR α -eGFP reporter mice, as detected by FACS analysis. **h–j**, The patterns of expression of **h**, **i**, and **j**, as assessed by FACS analysis of BaCl₂-injured muscle were similar to the patterns in **e**, **f**, and **g**, respectively, from uninjured muscle. **k–m**, Similarly, the patterns of expression of **k**, **l**, and **m**, as evaluated by FACS analysis of glycerol-injured muscle were comparable to the patterns in **e**, **f**, and **g**, respectively, from uninjured muscle.

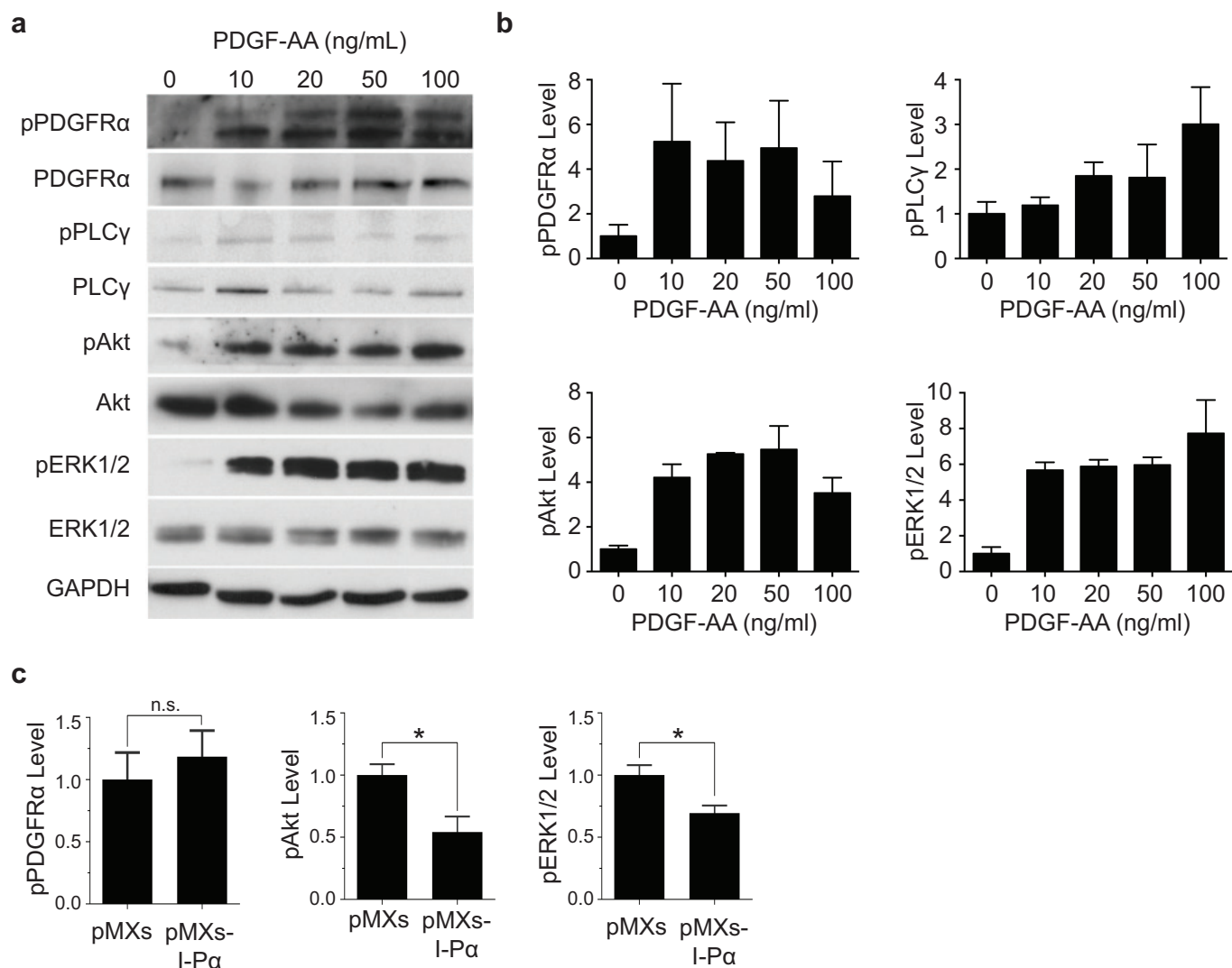


Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | In-PDGFR α transcript and protein structure.

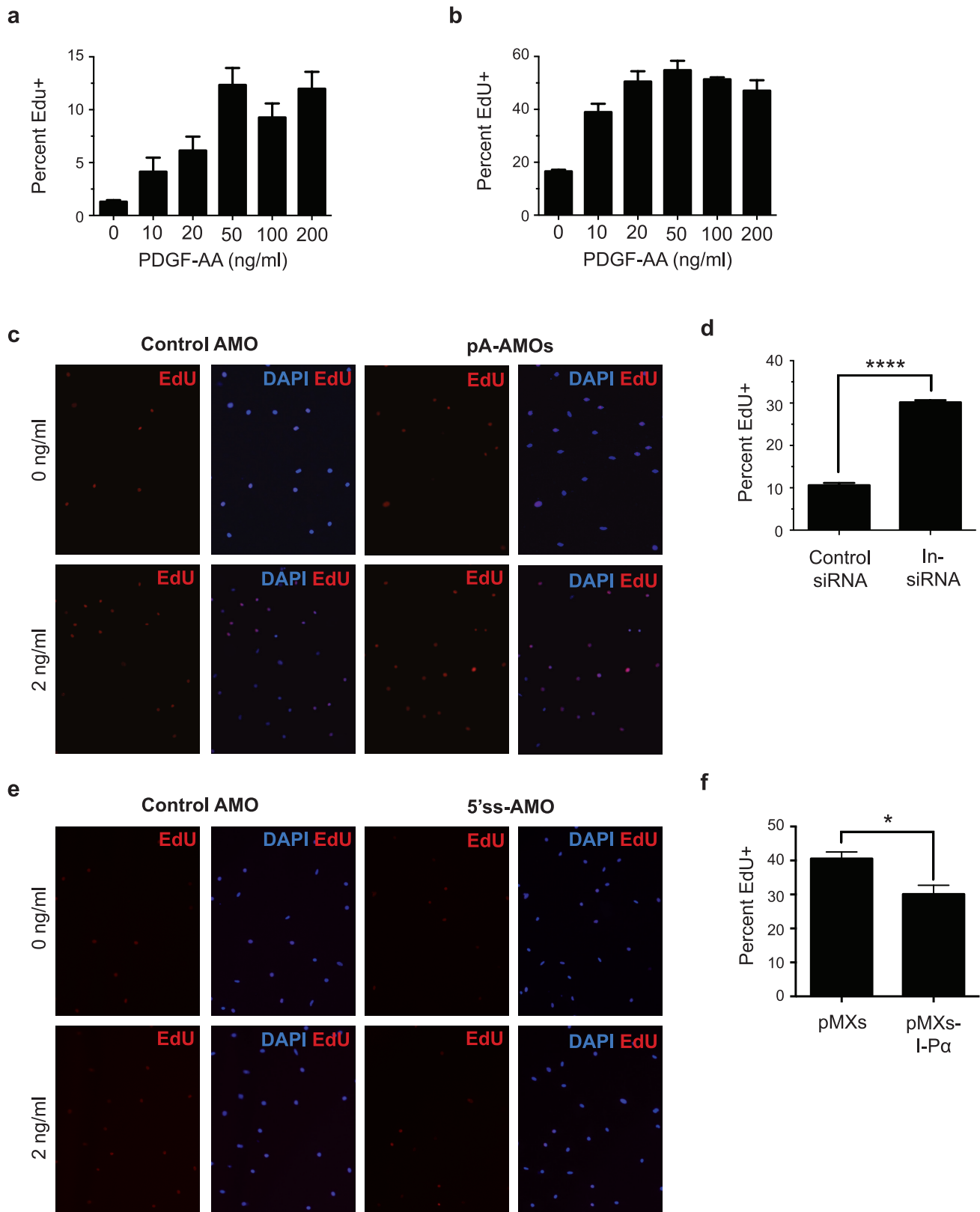
a, DNA sequencing of 3' RACE products illustrated in Fig. 1b confirms polyadenylation sites of highly expressed variants. **b**, Top, the fully-spliced FL-PDGFR α transcript contains 23 exons that encode the corresponding protein domains (bottom). CT, C-terminal region; JMR, juxtramembrane region; KD1, kinase domain 1; KD2, kinase domain 2; KI, interkinase domain; TM, transmembrane region. **c**, The fully spliced In-PDGFR α transcript contains 16 exons (top) that encode the protein domains (bottom). Red, portions of the transcript and protein that are unique to In-PDGFR α . **d**, Enlarged view of the genomic sequence that codes for the unique region of the In-PDGFR α protein. In FL-PDGFR α , this region is spliced out of the transcript. **e**, Map of the locations of amplicons used to assess levels of In-PDGFR α and FL-PDGFR α . Primers amplifying regions of exons 7–8, 11–12, and 15–16 are common to In-PDGFR α and FL-PDGFR α . Primers designated with 16i target the region canonically

referred to as intron 16. In FL-PDGFR α , this region is spliced out. In In-PDGFR α , this region becomes the 3' UTR. Therefore 16–16i and 16i–16i are specific for In-PDGFR α . Primers amplifying regions of exons 18–19, 21–22, exon 23 (23–23) and the 3' UTR (UTR–UTR) are specific to FL-PDGFR α . **f**, Levels of In-PDGFR α transcript relative to FL-PDGFR α transcript increase during regeneration. FAPs were collected from uninjured muscles (day 0) or at 3 days after injury and RNA levels were assessed via qPCR. Expression level is plotted as a ratio of In-PDGFR α to FL-PDGFR α normalized to day 0. For In-PDGFR α primers 16–16i and 16i–16i were averaged, whereas for FL-PDGFR α , primers 23–23 and UTR–UTR were averaged. For each time point, $n = 3$ biological replicates of pooled FAPs. Significance was calculated using an unpaired Student's t -test, error bars represent s.e.m. **g**, Western blot using the C-terminal PDGFR α antibody shows knockdown of FL-PDGFR α in response to the 5'ss-AMO.



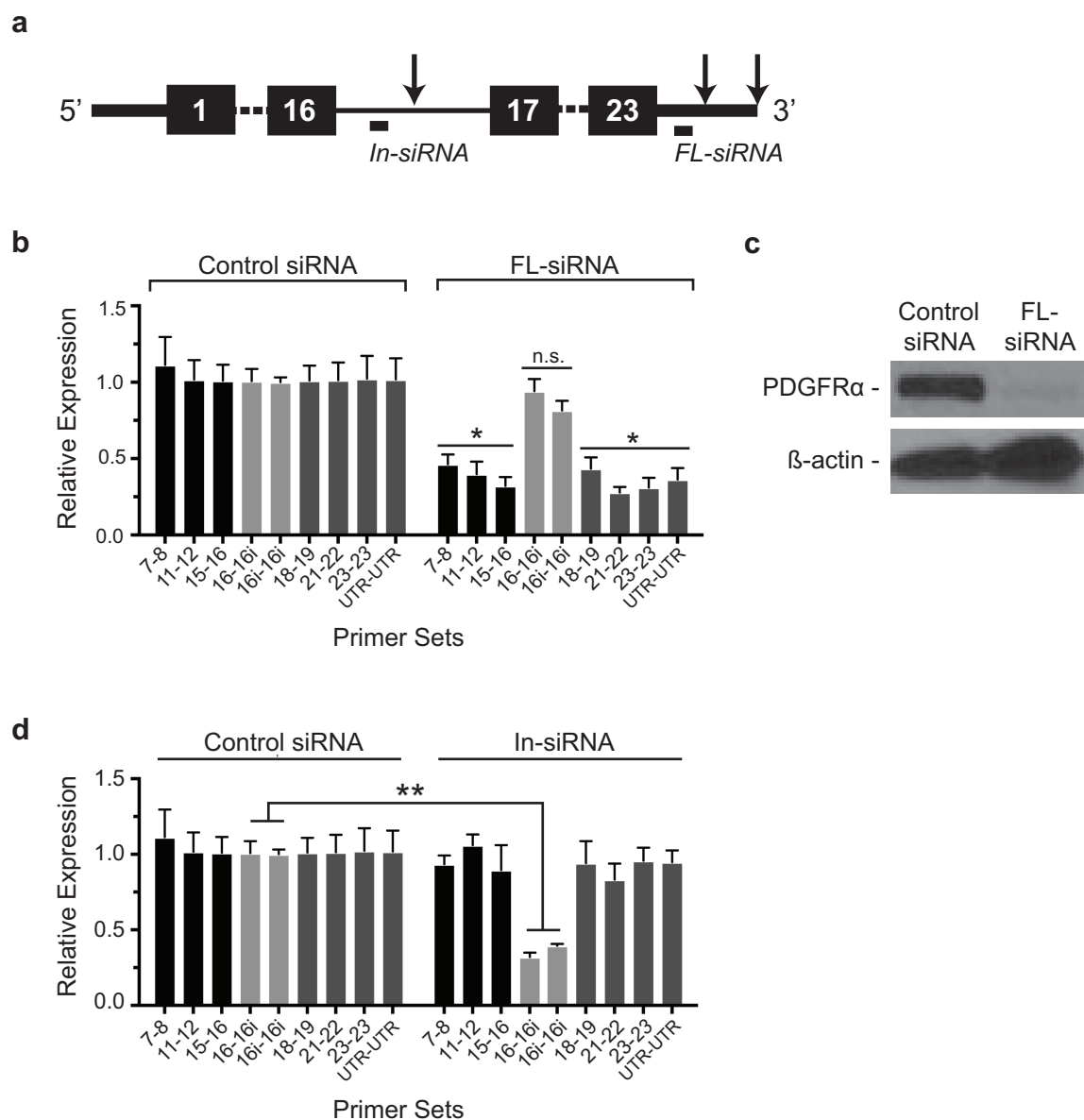
Extended Data Figure 3 | Phosphorylation of PDGFR α and downstream signalling components in FAPs in response to PDGF-AA stimulation is altered by changes in In-PDGFR α expression. a, Western blot showing phosphorylation of PDGFR α , PLC γ , Akt and ERK1/2 in response to PDGF-AA stimulation. **b**, Quantification of data in **a**. **c**, Viral

overexpression of In-PDGFR α in FAPs results in decreased signalling through Akt and ERK/2 whereas FL-PDGFR α levels remain constant. **b**, **c**, $n = 3$ biological replicates of pooled FAPs per condition except for the ERK1/2 condition in **b**, where $n = 4$. Error bars represent s.e.m. * $P < 0.05$; ** $P < 0.01$; unpaired Student t -tests.



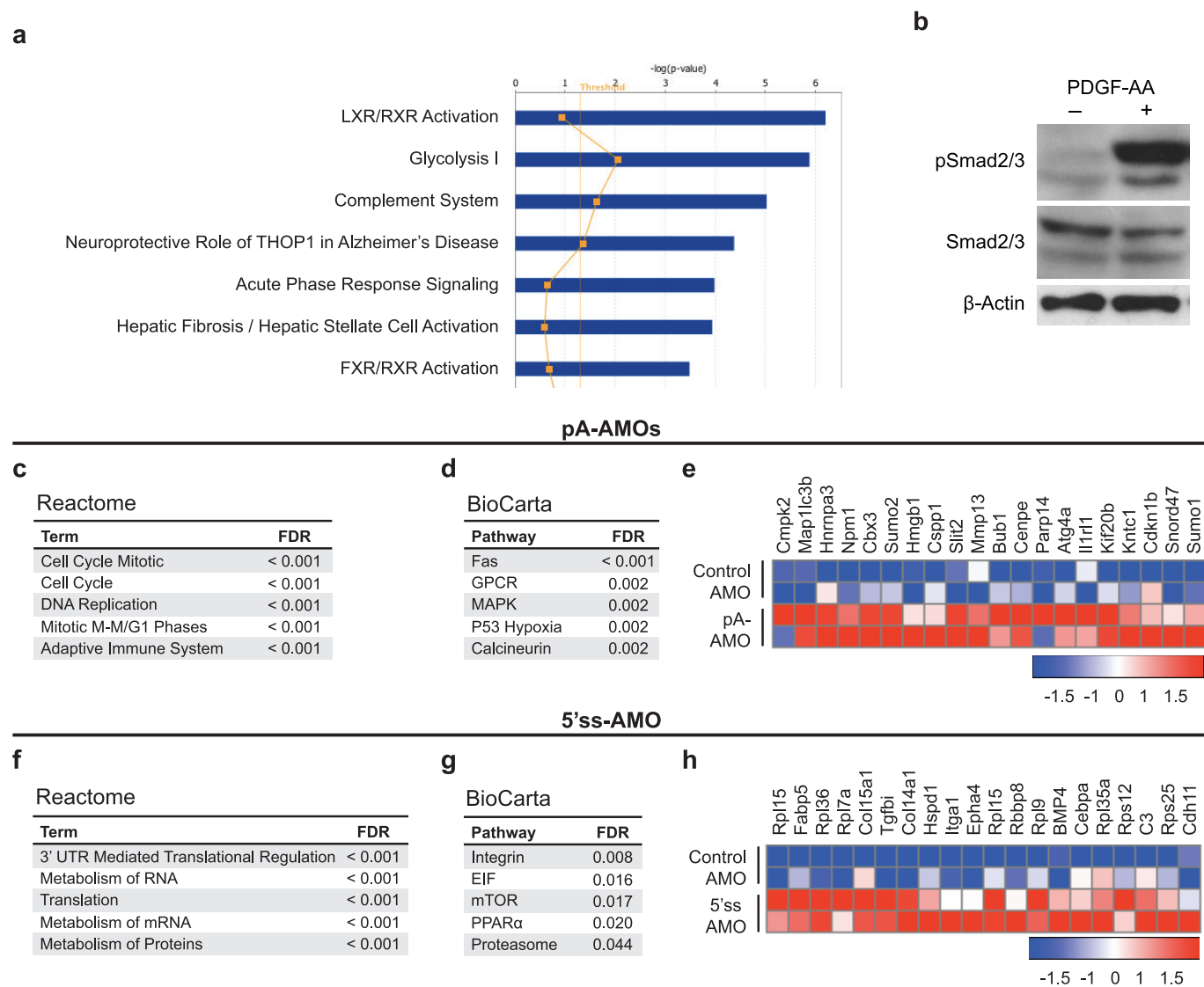
Extended Data Figure 4 | FAPs proliferate in response to PDGF stimulation. **a**, EdU incorporation in FAPs stimulated with indicated amounts of PDGF for 20 h and incubated in EdU for the final 4 h. **b**, EdU incorporation in FAPs incubated with indicated amounts of PDGF and EdU for 24 h. **c–f**, Knockdown of In-PDGFR α by pre-treatment of FAPs with pA-AMOs (**c**) or In-PDGFR α -specific siRNA (In-siRNA)

(**d**) increases FAP proliferation in response to PDGF-AA, whereas In-PDGFR α upregulation by 5'ss-AMO pre-treatment (**e**) or viral overexpression (**f**) decreases proliferation. For **a**, **b**, **d**, **f**, $n = 3$ biological replicates of pooled FAPs per condition. Error bars represent s.e.m. * $P < 0.05$; **** $P < 0.0001$; using unpaired Student t -tests.



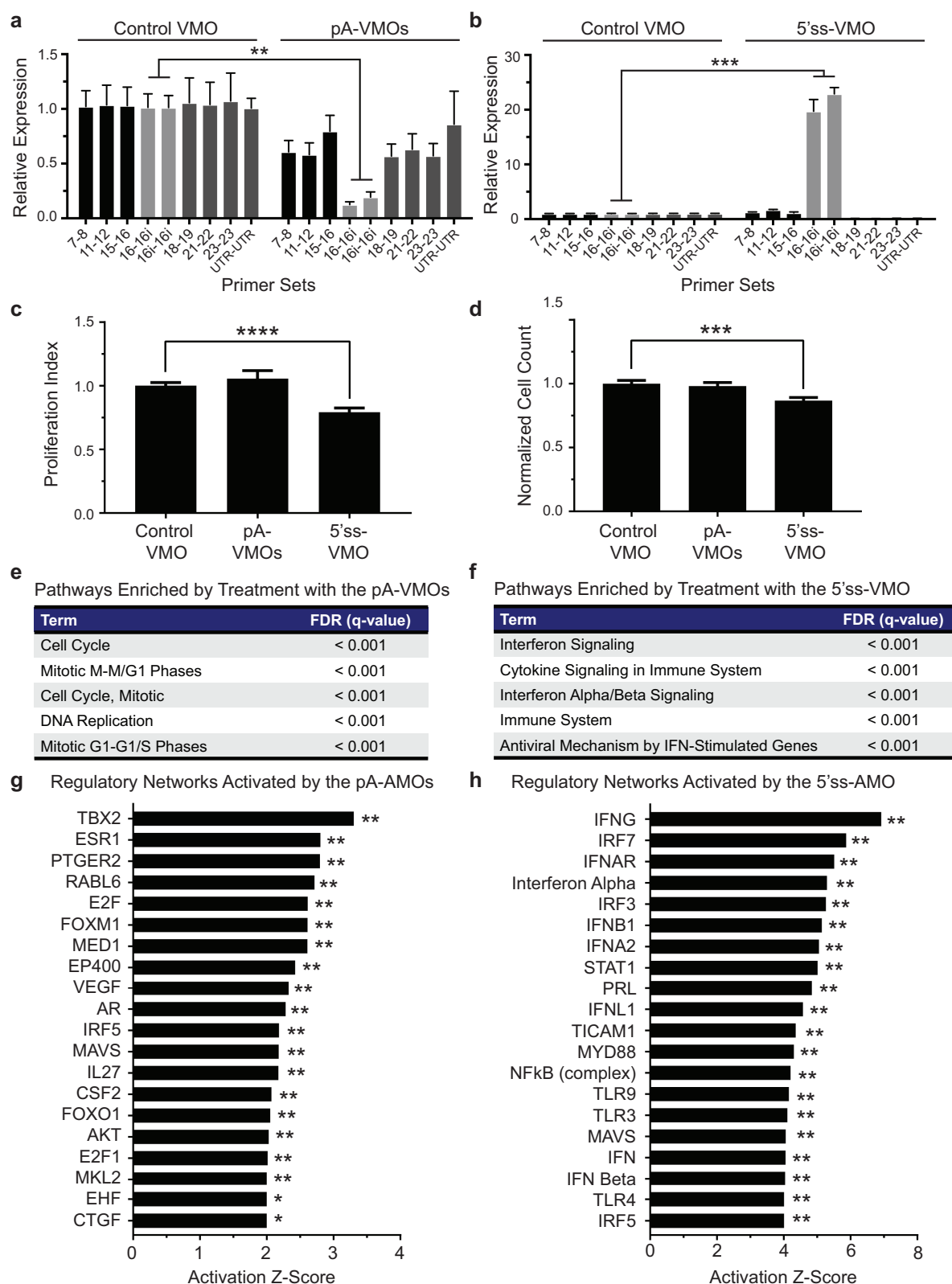
Extended Data Figure 5 | Knockdown of In-PDGFR α and FL-PDGFR α with siRNAs. **a**, Schematic of the FL-PDGFR α transcript showing the location of siRNAs designed to knockdown In-PDGFR α (In-siRNA) and FL-PDGFR α (FL-siRNA). Arrows designate relevant polyadenylation sites. **b–d**, FL-siRNA induces transcriptional knockdown of FL-PDGFR α but not In-PDGFR α (**b**) and protein knockdown of FL-PDGFR α (**c**).

At the same time, In-siRNA induces knockdown of In-PDGFR α (**d**). In **b** and **d**, $n = 3$ biological replicates of pooled FAPs per condition. The samples in **b**, and **d**, were all processed together so that the control condition in **b** is identical to the control condition in **d**. Error bars represent s.e.m. * $P < 0.05$, ** $P < 0.01$; unpaired Student t -tests.



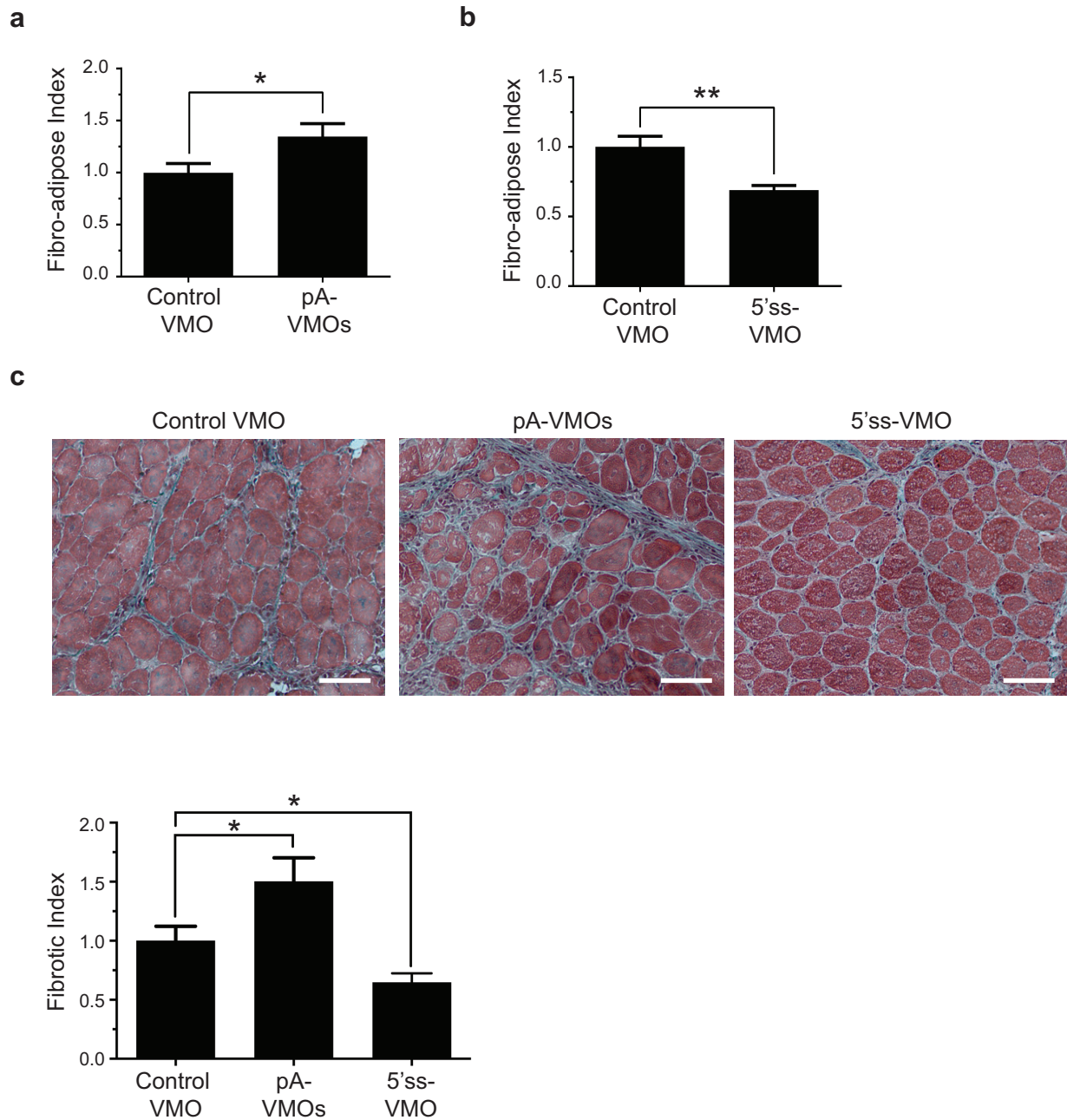
Extended Data Figure 6 | Assessment of molecular pathway changes in response to modulation of In-PDGFR α levels. **a**, Ingenuity pathway analysis of control-AMO-treated FAPs stimulated with PDGF-AA (50 ng ml⁻¹) show enrichment for genes associated with fibrosis compared with control cells. **b**, Western blot showing phosphorylation of SMAD2/3 in response to PDGF-AA stimulation (50 ng ml⁻¹). **c**, **d**, Gene set enrichment analysis (GSEA) of FAPs treated with the pA-AMOs compared to control-treated samples were analysed for enrichment of pathways in the Reactome (**c**) and BioCarta (**d**) databases^{30,31}. Top sets with a false

discovery rate of less than 5% are shown. **e**, The heat map displays a subset of genes that are upregulated with a false discovery rate of less than 0.1% in the pA-AMO- treated FAPs. The colours indicate the Z-score for expression normalized for each gene. Red, high expression; blue, low expression. **f**, **g**, Similarly, enrichment analyses of FAPs treated with the 5'ss-VMO compared to controls were performed using the Reactome (**f**) and BioCarta (**g**) databases^{30,31}. **h**, A heat map displaying top upregulated genes in 5'ss-VMO-treated FAPs with parameters as specified in **e** is shown.



Extended Data Figure 7 | Alteration of In-PDGFR α levels *in vivo* modulates FAP activity. **a, b**, FAPs treated *in vitro* with pA-VMOs showed a downregulation of In-PDGFR α ($n = 3$) (**a**), whereas those treated with the 5'ss-VMO exhibited an upregulation of In-PDGFR α ($n = 3$) (**b**). **c, d**, *In vivo* treatment with 5'ss-VMO decreases FAP proliferation ($n = 16$) (**c**) with a corresponding decrease in cell count ($n = 24$) (**d**). Treatment with pA-VMOs does not lead to a significant change in proliferation ($P = 0.42$). **e, f**, GSEA of FAPs collected from tibialis anterior muscles treated with pA-VMOs (**e**) or the 5'ss-VMO (**f**), compared to control-treated samples, were analysed for enrichment of pathways in

the Reactome database^{30,31}. Top sets with a false discovery rate of less than 5% are shown. **g, h**, Ingenuity pathway analysis of top regulators of gene expression in FAPs treated with pA-AMOs (**g**) or the 5'ss-VMO (**h**) compared to control treatment. The top hits are shown, excluding those with the 'molecular type' designated as a chemical or drug. For **a–d**, n represents biological replicates of pooled FAPs. * $P < 0.05$; ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$; unpaired Student t -tests. In **g, h**, two pooled FAP samples per condition were used with the overlap P -value calculated using the Fisher's exact test. Error bars represent s.e.m.



Extended Data Figure 8 | Alteration of In-PDGFR α levels *in vivo* modulates muscle fibrosis. a, b, Fibro-adipose index of muscles treated with pA-VMOs ($n=9$) (a) or the 5'ss VMO (control: $n=9$, 5'ss-VMO: $n=10$) (b). c, Representative images of Gomori-trichrome staining and quantification of fibrosis of BaCl₂-treated muscles from aged mice treated

with the control VMO ($n=10$), pA-VMOs ($n=9$) or the 5'ss-VMO ($n=10$). For a–c, n represents biological replicates consisting of tibialis anterior muscles isolated from individual mice. Scale bars, 100 μ m. Error bars represent s.e.m. * $P < 0.05$; ** $P < 0.01$; unpaired Student's t test.

Extended Data Table 1 | *Pdgfra* gene-specific primers used in 3' RACE

Location	Sequence
Exon 2	CTATGGGGACCTCCCACCAGGTCTTTCT
Exon 10	GTCCTTCGCCAAAGTGGAAGAGACCATC
Exon 16	AACGGCGACTACATGGACATGAAGCAAG
3'-UTR (precedes both polyA sites)	CCTCTCCGGACCTCTGAAGAGACCACTC

Extended Data Table 2 | Primers for *Pdgfra* variant quantification

Name	Forward Sequence	Reverse Sequence
7-8	CTCACCGAGATCACCCTGA	CCTTAGCCCGGATCAGCTTT
11-12	TGTTGGTGCTGTTGGTGATT	CGTGGTTTCTGCTTCAAAT
15-16	TGAATCCTGCAGACGAGAGC	ATGTCCATGTAGTCGCCGTT
16-16i	GGCCAGCCTCCTACAAGAAG	TCCCTCTGAGCAGCAAGTTA
16i-16i	AAAAGTGCCCATGCTCATTC	GCTTGGCAGAGCTACCTGTC
18-19	CTGGCCAGAGACATCATGCA	CATCCACTTCACAGGCAGGA
21-22	TGTTGCCGGGACAATACAAGA	TATCAGAGTCCACCCGCATG
23-23	GAGAGAGGACGAGACCATCG	CACCAGGTCCGAGGAATCTA
UTR-UTR	AGAGGACTTGGGTGATGTGG	TCATCCACACAGGCTTACCA

Microcins mediate competition among Enterobacteriaceae in the inflamed gut

Martina Sassone-Corsi^{1,2}, Sean-Paul Nuccio^{1,2}, Henry Liu¹, Dulcemaria Hernandez¹, Christine T. Vu¹, Amy A. Takahashi¹, Robert A. Edwards^{2,3} & Manuela Raffatellu^{1,2}

The Enterobacteriaceae are a family of Gram-negative bacteria that include commensal organisms as well as primary and opportunistic pathogens that are among the leading causes of morbidity and mortality worldwide. Although Enterobacteriaceae often comprise less than 1% of a healthy intestine's microbiota¹, some of these organisms can bloom in the inflamed gut^{2–5}; expansion of enterobacteria is a hallmark of microbial imbalance known as dysbiosis⁶. Microcins are small secreted proteins that possess antimicrobial activity *in vitro*^{7,8}, but whose role *in vivo* has been unclear. Here we demonstrate that microcins enable the probiotic bacterium *Escherichia coli* Nissle 1917 (EcN) to limit the expansion of competing Enterobacteriaceae (including pathogens and pathobionts) during intestinal inflammation. Microcin-producing EcN limits the growth of competitors in the inflamed intestine, including commensal *E. coli*, adherent-invasive *E. coli* and the related pathogen *Salmonella enterica*. Moreover, only therapeutic administration of the wild-type, microcin-producing EcN to mice previously infected with *S. enterica* substantially reduced intestinal colonization by the pathogen. Our work provides the first evidence that microcins mediate inter- and intraspecies competition among the Enterobacteriaceae in the inflamed gut. Moreover, we show that microcins can act as narrow-spectrum therapeutics to inhibit enteric pathogens and reduce enterobacterial blooms.

Broad-spectrum antibiotics are largely unsuccessful in treating enterobacterial gut infections. Efforts to identify novel therapeutic treatments can be aided by identifying the mechanisms employed by beneficial microbes to mediate colonization resistance to pathogens^{9,10}. Reports of microcin-producing Enterobacteriaceae in human stool led to suggestions that microcins contribute to gut microbial ecology¹¹. We therefore sought to identify an *in vivo* role for microcins by studying EcN^{12,13} and its interplay with related competitors in the gut. Microcins can exhibit potent antibacterial activity *in vitro*, similar to bacteriocins produced by Gram-positive bacteria^{14–16}. In contrast to bacteriocins¹⁵, however, a compelling *in vivo* role for microcins has not previously been demonstrated¹⁷.

The EcN microcin gene cluster resides on genomic island I (Extended Data Fig. 1a and Supplementary Table 1) and encodes for microcin M (*mcmA*), microcin H47 (*mchB*) and their cognate immunity genes (*mcmI* and *mchI*, respectively). To understand whether the microcins of EcN promote colonization of the healthy intestine, we intragastrically inoculated specific-pathogen-free (SPF) mice with an equal mixture of a mouse-commensal *E. coli* (cEc) strain¹⁸ and either wild-type EcN or a mutant unable to secrete both microcins (EcN *mchDEF*; all strains are described in Supplementary Table 3)¹⁴. Although all strains poorly and transiently colonized the healthy intestine (enterobacteria typically colonize at low levels in this setting), cEc colonization was not significantly impacted by either wild-type EcN or EcN *mchDEF*, and cEc outcompeted both strains to a similar extent (Extended Data Fig. 1b–g).

To determine whether low colonization masked a microcin phenotype, we performed similar studies in germ-free mice and in streptomycin-treated SPF mice. We administered cEc alone or in competition with wild-type EcN, EcN *mchDEF*, or EcN *mcmA mchB*, a mutant that lacks both microcin genes. All strains colonized the gut to high levels in these models, but cEc colonization was again unaffected by competition with EcN wild-type or microcin mutants (Fig. 1a, b). Moreover, there were no significant differences in colonization among EcN strains (Fig. 1c, d), and cEc outcompeted EcN wild-type and microcin mutants to a similar extent (Fig. 1e, f). Together, these results indicate that the microcins of EcN do not provide a competitive advantage under homeostatic conditions (Fig. 1g).

Prior work has shown that the microcins of EcN possess antibacterial activity in iron-limited medium¹⁴. Moreover, a subset of microcins are post-translationally modified at the C terminus with a siderophore moiety^{19–21} and EcN microcins are thought to be modified by a similar mechanism^{14,21}. In response to iron starvation, some commensals and pathogens synthesize and secrete siderophores to acquire this essential metal nutrient²². Bacterial uptake of iron-bound siderophores occurs via specific membrane receptors¹⁴. Microcins conjugated to siderophores are also synthesized during iron starvation and can target non-immune bacteria *in vitro* through cognate siderophore receptors^{14,16}, a mechanism described as a ‘Trojan horse’ mode of action⁷. Incorporating these observations, we proposed that microcins contribute to competition among enterobacteria in the inflamed gut, where iron is limited²³.

We confirmed that EcN microcin genes (*mcmA*, *mchB*) are expressed in iron-limited medium, and not in iron-rich medium (Fig. 1h). To identify a role for microcins during intestinal inflammation, we used the dextran sulphate sodium (DSS)-treated mouse model of colitis (Fig. 2a). First, we inoculated DSS-treated SPF mice intragastrically with an equal mixture of cEc and either wild-type EcN, EcN *mchDEF*, or EcN *mcmA mchB*. Although cEc initially outcompeted (~10-fold) wild-type EcN, both strains colonized the intestine equally well by 5 days after inoculation. By contrast, cEc greatly outcompeted the microcin mutants (up to 10,000-fold; Fig. 2b and Extended Data Fig. 2a). Whereas intestinal colonization of cEc was similar when alone or in competition with wild-type EcN or a microcin mutant at day 1–3 following inoculation, only wild-type EcN significantly reduced cEc colonization at later time points (Fig. 2c and Extended Data Fig. 2b, h). Consistent with this, the *in vitro* killing of cEc was only observed in iron-limited medium when competing with wild-type EcN, not the microcin mutants (Extended Data Fig. 3a–d). As the EcN microcin mutants still express cognate microcin immunity proteins, their competitive growth *in vitro* was not impaired (Extended Data Fig. 3c–g) until the immunity genes were deleted (Extended Data Fig. 3h, i). Nevertheless, gut colonization of microcin mutants was similarly reduced in competition with cEc (Fig. 2d and Extended Data Fig. 2c, i), but complementation of EcN *mchDEF* with *mchDEF* (EcN *mchDEF* pWSK29::*mchDEF*) restored colonization

¹Department of Microbiology and Molecular Genetics, University of California Irvine, Irvine, California 92697, USA. ²Institute for Immunology, University of California Irvine, Irvine, California 92697, USA. ³Department of Pathology and Laboratory Medicine, University of California, Irvine, California 92697, USA.

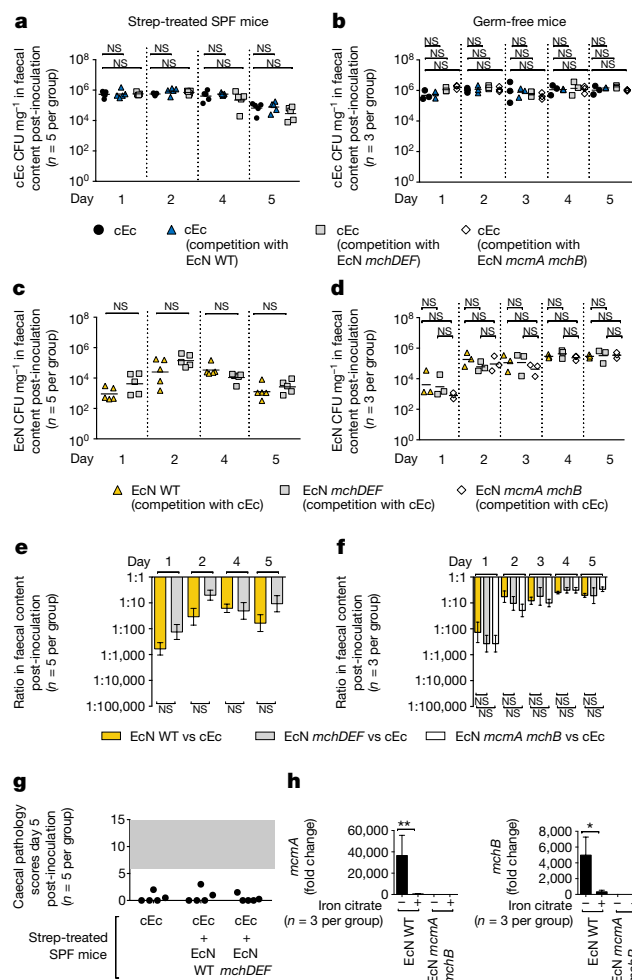


Figure 1 | Microcins do not promote bacterial competition in the absence of intestinal inflammation. **a–g**, SPF streptomycin-treated (at 1 day prior to infection) mice (**a**, **c**, **e**, **g**) or germ-free mice (**b**, **d**, **f**) were intragastrically inoculated with the indicated *E. coli* strains. **a–d**, Colony-forming units (CFU) mg^{-1} in the faeces of SPF (**a**, **c**) or germ-free (**b**, **d**) mice on days 1–5 following inoculation. **e**, **f**, Ratio (EcN:cEc) of panels **a** and **c** (in **e**) or panels **b** and **d** (in **f**); bars represent the geometric mean \pm s.e.m. **g**, Caecal histopathology scores for mice in panels **a** and **c**. **a–d**, **g**, Each symbol represents an individual mouse, bars represent the geometric mean. **h**, Levels of *mcmA* (microcin M) and *mchB* (microcin H47) mRNA in iron-rich and iron-limited medium, normalized to the *gapA* housekeeping gene; bars represent the geometric mean from three independent experiments \pm s.e.m. WT, wild type. * $P < 0.05$, ** $P < 0.01$; NS, not significant. Unpaired Student's *t*-test was used for statistical comparisons in panel **h**. Mann–Whitney–Wilcoxon for all other comparisons; *P* values are presented in Supplementary Table 7.

(Fig. 2b–d). As cEc was isolated from the murine gut, cEc is plausibly better adapted to growth in this environment than is EcN, a human isolate. Microcins thus give EcN a means by which to compete.

Colonization of microcin mutants was not impaired in DSS-treated mice when strains were administered alone or in competition with wild-type EcN, in contrast to competition with cEc (Fig. 3a–c and Extended Data Fig. 4a–d). Microcin mutants exhibited a 5–10-fold advantage over wild-type EcN at day 5 after inoculation (Fig. 3c). This observation may result from the mutants not having to expend energy on microcin production, yet still being immune to microcin activity; relatedly, wild-type EcN rescued the colonization of microcin mutants when co-administered with cEc in DSS-treated mice (Extended Data Fig. 4e–h). Microcin mutants are thus not intrinsically attenuated. Moreover, our findings indicate that microcins enable EcN to limit the expansion of a related commensal in the inflamed gut (Extended

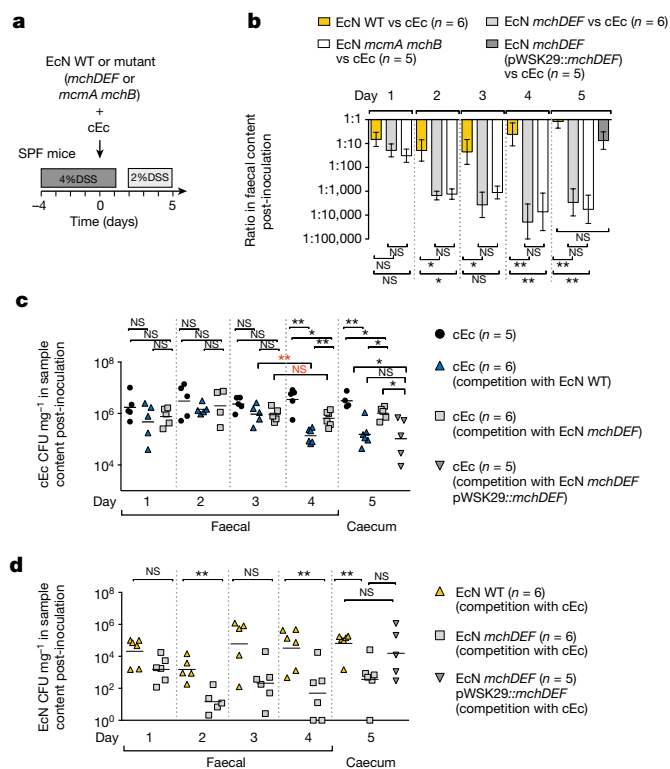


Figure 2 | Microcins enable *E. coli* Nissle to limit the expansion of a commensal *E. coli* in the inflamed gut. **a**, Experimental design of **b–d**. **b–d**, Ratio (EcN:commensal) (**b**) or CFU mg^{-1} (**c**, **d**) sample at days 1–5 after intragastric inoculation of indicated *E. coli* strains. Each symbol represents an individual mouse (**c**, **d**). Bars represent the geometric mean \pm s.e.m. (**b**) or the geometric mean (**c**, **d**) for two experiments; * $P < 0.05$, ** $P < 0.01$; NS, not significant. Statistics in black compare same time point; red compare different time points. Mann–Whitney–Wilcoxon was used for all statistical comparisons; *P* values are presented in Supplementary Table 7.

Data Fig. 2d–g), where Enterobacteriaceae thrive but also compete with each other for space and limited nutrients (niche competition).

We next tested whether iron supplementation during colitis inhibited microcin activity. Mice fed an iron-rich diet three weeks before and during DSS treatment developed colitis, yet wild-type EcN did not reduce colonization of cEc. Instead, both wild-type EcN and EcN *mchDEF* were outcompeted by cEc and exhibited similar levels of gut colonization (Extended data Fig. 5a–e). Thus, if iron is supplemented, microcins do not limit the expansion of a competitor in the inflamed gut, probably because microcins, siderophores and siderophore receptors (that is, microcin receptors) are only expressed during iron starvation^{14,23}. These observations also suggest that, in contrast to broad-spectrum antibiotics, siderophore-linked microcins would have a limited effect on the gut microbiota as they only target strains that acquire those siderophores.

As microcins mediated competition between EcN and cEc in the DSS–colitis model, we analysed the intestinal microbiota in this setting. DSS treatment reduced the gut microbiota's diversity (Extended Data Fig. 5f) and the subsequent administration of wild-type EcN or EcN *mchDEF* yielded similar shifts in microbial communities (Extended Data Fig. 5f, g and Supplementary Table 2). These results indicate that, at least in our DSS-treated mice, EcN microcins did not substantially change the gut microbiota, which is consistent with their predicted narrow spectrum.

In addition to some commensal Enterobacteriaceae, some enterobacterial pathogens thrive in the inflamed gut³. We therefore investigated microcin activity against the diarrheal pathogen *Salmonella enterica* serovar Typhimurium (STm). We first established that wild-type

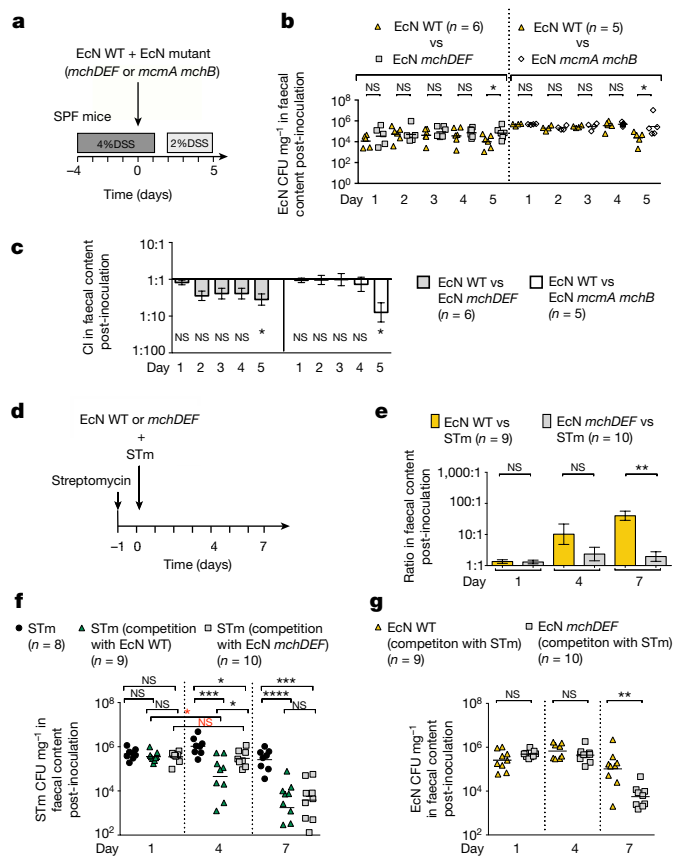


Figure 3 | *E. coli* Nissle utilizes microcins to outcompete an enteric pathogen when the strains are co-administered. **a**, The experimental design for **b** and **c**. **b**, **c**, CFU mg⁻¹ of faeces (**b**) or competitive index (CI; EcN wild-type over mutant) (**c**) at days 1–5 following intragastric inoculation of the indicated *E. coli* strains. **d**, Co-administration design for **e**–**g**. **e**–**g**, Ratio (EcN:STm) (**e**) or CFU mg⁻¹ faeces (**f**, **g**) at designated time points after intragastric inoculation of the indicated strains. **c**, **e**, Bars represent the geometric mean ± s.e.m. **b**, **f**, **g**, Each symbol represents one mouse; bars represent the geometric mean from 2 independent experiments. **P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001; NS, not significant. Statistics in black compare the same time point; those in red compare different time points. Mann–Whitney–Wilcoxon was used for all statistical comparisons; *P* values are presented in Supplementary Table 7.

EcN (but not EcN *mchDEF* or EcN *mcmA mchB*) only killed STm in iron-limited medium (Extended Data Fig. 6a–e). When streptomycin-treated mice (a *Salmonella* colitis model) were co-administered STm and either wild-type EcN or EcN *mchDEF*, only wild-type EcN outcompeted STm beyond day 4 after infection (Fig. 3d, e). Although all strains initially colonized the gut to similar levels, reduction of STm colonization was significantly enhanced by microcin-producing EcN (Fig. 3f). By day 7, wild-type EcN and EcN *mchDEF* had similarly reduced STm colonization, probably because EcN has other mechanisms to outcompete STm, including a greater ability to acquire iron²³. Nevertheless, in agreement with our cEc results (Fig. 2a–d), gut colonization of EcN *mchDEF* was drastically reduced in competition with STm 7 days after co-administration (Fig. 3g). Moreover, wild-type EcN's competitive advantage over STm was again only observed during intestinal inflammation (Extended Data Fig. 7a, b). Relatedly, wild-type EcN and EcN *mchDEF* exhibited similar levels of intestinal colonization and no competitive advantages when competed with STm *invA spiB*, a mutant in both type-three secretion systems that is unable to trigger gut inflammation as it cannot invade the intestinal epithelium or replicate intracellularly (Extended Data Fig. 7c–e).

Consistent with our results from STm and cEc experiments, microcins impaired growth of adherent invasive *E. coli* (AIEC), a pathobiont

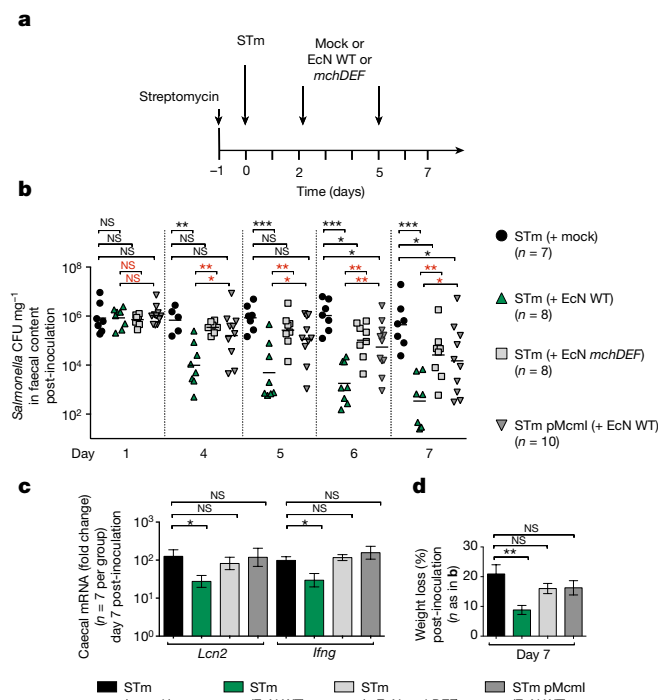


Figure 4 | Microcins therapeutically reduce colonization of an enteric pathogen in the inflamed gut. **a**, Therapeutic design for **b**–**d** and Extended Data Fig. 9. **b**, *Salmonella* CFU mg⁻¹ in mouse faeces on days 1 and 4–7 following infection. **c**, Caecal levels of indicated mRNAs 7 days after infection. **d**, Weight loss of mice infected with indicated strains 7 days after infection. Each symbol represents one mouse (**b**). Bars represent the geometric mean from two independent experiments (**b**), geometric mean ± s.e.m. (**c**) or mean ± s.e.m. (**d**). **P* < 0.05, ***P* < 0.01, ****P* < 0.001; NS, not significant. Statistics in black compare to STm (+ mock; vehicle control); those in red compare to STm (+ wild-type EcN). Mann–Whitney–Wilcoxon test was used for all statistical comparisons; *P* values are presented in Supplementary Table 7.

frequently isolated from patients with Crohn's disease. Wild-type EcN (but not EcN *mchDEF*) was able to reduce AIEC growth *in vitro* (Extended Data Fig. 6f–i) and in the DSS–colitis model (Extended Data Fig. 8). Although AIEC and wild-type EcN colonized the inflamed gut to similar levels, AIEC significantly outcompeted EcN *mchDEF* in this environment (Extended Data Fig. 8). Collectively, our data indicate that EcN produces microcins to limit growth of competing enterics in the inflamed gut.

We next investigated whether microcins could therapeutically treat an active enterobacterial infection by administering wild-type EcN or EcN *mchDEF* to mice previously infected with STm (Fig. 4a). Mice receiving wild-type EcN at days 2 and 5 after infection with STm showed substantially reduced levels of STm gut colonization (average 1,300-fold; up to 100,000-fold), inflammation and weight loss (Fig. 4b–d and Extended Data Fig. 9). This phenotype largely depended on microcins, as EcN *mchDEF* only minimally reduced STm intestinal burden (approximately 17-fold) and did not reduce inflammation or weight loss (Fig. 4b–d and Extended Data Fig. 9).

To understand the therapeutic phenotype better, we assayed the antibacterial activity of each microcin. Although both M and H47 microcins showed antibacterial activity against AIEC, only M was effective against STm and cEc (Extended Data Fig. 10a–c); the H47-resistance mechanism is unknown as neither strain carries the cognate immunity gene *mchI*. Accordingly, STm engineered to express *mchI* (STm pMchI) was not rescued in competition with wild-type EcN, whereas STm engineered to express *mcmI* (STm pMcmI) was rescued (Extended Data Fig. 10d, e). Notably, the therapeutic administration of EcN only

minimally impaired colonization of the microcin-resistant STm pMcmI (Fig. 4b and Extended Data Fig. 9). Furthermore, inflammation, weight loss and STm colonization were comparable when either EcN *mchDEF* or STm pMcmI were used in the therapeutic experiment (Fig. 4b–d and Extended Data Fig. 9).

Together, our data provide *in vivo* evidence that microcins mediate competition among Enterobacteriaceae in the inflamed gut. We show that the probiotic bacterium EcN employs microcins to compete with related species, colonize the inflamed gut, and therapeutically displace an enteric pathogen from its niche. The microcins produced by EcN target microbes that express particular siderophore receptors and do not severely affect the microbiota. Although siderophore-receptor mutants are resistant to microcins¹⁴ (Extended Data Fig. 10f–h), they are themselves attenuated in the inflamed intestine²⁴ and thus unlikely to be selected for. As microcins were particularly effective upon therapeutic EcN administration to pathogen-infected, inflamed mice, these or similar compounds (for example, siderophore-conjugated antibiotics²⁵) could conceivably be used as a targeted strategy to treat enterobacterial colitis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 January; accepted 24 October 2016.

Published online 31 October 2016.

- Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14**, 20–32 (2016).
- Winter, S. E. & Bäuml, A. J. Why related bacterial species bloom simultaneously in the gut: principles underlying the ‘Like will to like’ concept. *Cell. Microbiol.* **16**, 179–184 (2014).
- Stecher, B. The roles of inflammation, nutrient availability and the commensal microbiota in enteric pathogen infection. *Microbiol. Spectr.* **3**, MBP-0008-2014 (2015).
- Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96–99 (2013).
- Winter, S. E. *et al.* Host-derived nitrate boosts growth of *E. coli* in the inflamed gut. *Science* **339**, 708–711 (2013).
- Winter, S. E., Lopez, C. A. & Bäuml, A. J. The dynamics of gut-associated microbial communities during inflammation. *EMBO Rep.* **14**, 319–327 (2013).
- Rebuffat, S. Microcins in action: amazing defence strategies of Enterobacteria. *Biochem. Soc. Trans.* **40**, 1456–1462 (2012).
- Yang, S. C., Lin, C. H., Sung, C. T. & Fang, J. Y. Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. *Front. Microbiol.* **5**, 241 (2014).
- Kamada, N., Chen, G. Y., Inohara, N. & Núñez, G. Control of pathogens and pathobionts by the gut microbiota. *Nat. Immunol.* **14**, 685–690 (2013).
- Buffie, C. G. & Pamer, E. G. Microbiota-mediated colonization resistance against intestinal pathogens. *Nat. Rev. Immunol.* **13**, 790–801 (2013).
- Baquero, F. & Moreno, F. The microcins. *FEMS Microbiol. Lett.* **23**, 117–124 (1978).
- Behnsen, J., Deriu, E., Sassone-Corsi, M. & Raffatellu, M. Probiotics: properties, examples, and specific applications. *Cold Spring Harb. Perspect. Med.* **3**, a010074 (2013).
- Jacobi, C. A. & Malfertheiner, P. *Escherichia coli* Nissle 1917 (Mutaflor): new insights into an old probiotic bacterium. *Dig. Dis.* **29**, 600–607 (2011).
- Patzer, S. I., Baquero, M. R., Bravo, D., Moreno, F. & Hantke, K. The colicin G, H and X determinants encode microcins M and H47, which might utilize the catecholate siderophore receptors FepA, Cir, Fiu and IroN. *Microbiology* **149**, 2557–2570 (2003).
- Kommineni, S. *et al.* Bacteriocin production augments niche competition by enterococci in the mammalian gastrointestinal tract. *Nature* **526**, 719–722 (2015).
- Nolan, E. M. & Walsh, C. T. Investigations of the MceJ-catalyzed posttranslational modification of the microcin E492 C-terminus: linkage of ribosomal and nonribosomal peptides to form ‘Trojan horse’ antibiotics. *Biochemistry* **47**, 9289–9299 (2008).
- Sassone-Corsi, M. & Raffatellu, M. No vacancy: how beneficial microbes cooperate with immunity to provide colonization resistance to pathogens. *J. Immunol.* **194**, 4081–4087 (2015).
- Behnsen, J. *et al.* The cytokine IL-22 promotes pathogen colonization by suppressing related commensal bacteria. *Immunity* **40**, 262–273 (2014).
- Duquesne, S., Destoumieux-Garzon, D., Peduzzi, J. & Rebuffat, S. Microcins, gene-encoded antibacterial peptides from enterobacteria. *Nat. Prod. Rep.* **24**, 708–734 (2007).
- Rebuffat, S., Blond, A., Destoumieux-Garzon, D., Goulard, C. & Peduzzi, J. Microcin J25, from the macrocyclic to the lasso structure: implications for biosynthetic, evolutionary and biotechnological perspectives. *Curr. Protein Pept. Sci.* **5**, 383–391 (2004).
- Vassiliadis, G., Destoumieux-Garzon, D., Lombard, C., Rebuffat, S. & Peduzzi, J. Isolation and characterization of two members of the siderophore-microcin family, microcins M and H47. *Antimicrob. Agents Chemother.* **54**, 288–297 (2010).
- Fischbach, M. A., Lin, H., Liu, D. R. & Walsh, C. T. How pathogenic bacteria evade mammalian sabotage in the battle for iron. *Nat. Chem. Biol.* **2**, 132–138 (2006).
- Deriu, E. *et al.* Probiotic bacteria reduce *Salmonella* typhimurium intestinal colonization by competing for iron. *Cell Host Microbe* **14**, 26–37 (2013).
- Raffatellu, M. *et al.* Lipocalin-2 resistance confers an advantage to *Salmonella enterica* serotype Typhimurium for growth and survival in the inflamed intestine. *Cell Host Microbe* **5**, 476–486 (2009).
- Zheng, T. & Nolan, E. M. Enterobactin-mediated delivery of β -lactam antibiotics enhances antibacterial activity against pathogenic *Escherichia coli*. *J. Am. Chem. Soc.* **136**, 9677–9691 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to acknowledge E. Nolan for critically reading the manuscript, J. Behnsen for setting up the Raffatellu laboratory germ-free mouse facility, and M. Valeri, A. Perez-Lopez, V. Diaz-Ochoa and E. Hoover for contributing to the maintenance of the germ-free facility. We would also like to thank W. Zhu and S. Winter for providing us with the protocol for cleaning up DNA stool samples from DSS, and M. Rolston at the UC Davis Host-Microbe Systems Biology Core for processing samples for Illumina MiSeq Analysis. We would like to acknowledge the students of the 2015 Summer course ‘Frontiers in Host-Microbe Interactions’, Marine Biology Laboratory, who helped with the generation of one dataset for the manuscript. Work in the laboratory of M.R. is supported by Public Health Service Grants AI083663, AI126277, AI101784, AI114625, AI105374, and DK058057. M.R. holds an Investigator in the Pathogenesis of Infectious Disease Award from the Burroughs Wellcome Fund.

Author Contributions M.S.-C. performed most bacterial growth assays and all animal experiments, and analysed the results. S.-P.N. analysed the microbiota data. R.A.E. scored histological sections. H.L., D.H., C.T.V., and A.A.T. assisted with mutant construction, bacterial growth assays, and animal experiments. M.S.-C. and M.R. were responsible for the overall study design. M.S.-C., S.-P.N. and M.R. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.R. (manuelar@uci.edu).

METHODS

Bacterial strains, culture conditions and chemicals. Strains of the probiotic *E. coli* Nissle 1917 (EcN; Mutaflor, DSM 6601), the mouse commensal *E. coli* (cEc), the adherent invasive *E. coli* (AIEC), and the pathogen *S. enterica* serovar Typhimurium (STm) used in this study are listed in Supplementary Table 3. All plasmids used in this study are listed in Supplementary Table 4. EcN was provided by Ardeypharm GmbH. The commensal *E. coli* strain was isolated from mice in our vivarium and does not appear to encode for known antibacterial exoproducts. The AIEC strain is a human isolate from a patient with Crohn's disease (isolate NRG857c O83:H1)²⁶ and was provided by A. Torres. The STm strain background we employed (IR715) is a fully virulent, nalidixic acid-resistant derivative of STm wild-type isolate ATCC 14028 and does not encode for known antibacterial exoproducts. All strains were routinely grown aerobically in Luria–Bertani (LB) broth (10 g l⁻¹ tryptone, 5 g l⁻¹ yeast extract, 10 g l⁻¹ NaCl) or on LB agar plates at 37 °C. When indicated, antibiotics were added to the medium at the following concentrations: 0.03 mg ml⁻¹ chloramphenicol (Cm); 0.1 mg ml⁻¹ carbenicillin (Carb); 0.05 mg ml⁻¹ kanamycin (Kan); 0.01 mg ml⁻¹ tetracycline (Tet). For animal infections or bacterial administration, all strains were grown in LB medium aerobically at 37 °C overnight. For *in vitro* growth assays, strains were grown in iron-limiting conditions (nutrient broth supplemented with 0.2 mM 2,2'-dipyridyl dissolved in ethanol; Sigma) aerobically at 37 °C overnight. Restriction enzymes and Phusion High Fidelity DNA Polymerase were purchased from New England Biolabs. Oligonucleotides were synthesized by Fisher Scientific and are listed in Supplementary Tables 5 and 6.

Generation of bacterial mutants. Mutants in EcN and STm were constructed using the lambda red recombinase system²⁷. In brief, primers (Supplementary Table 5) homologous to sequences flanking the 5' and 3' ends of the target regions were designed (H1 and H2 primers, respectively; Supplementary Table 5) and were used to replace the selected genes with a chloramphenicol- (derived from pKD3), kanamycin- (derived from pKD4) or tetracycline-resistance cassette (Supplementary Table 4). Strain names for the mutants are listed in Supplementary Table 3. To confirm integration of the resistance cassette and deletion of the target, mutant strains and wild-type controls were each assayed using three PCR amplifications (5' end, 3' end, deleted target) to validate mutations. Primers (Supplementary Table 5) that flank the target sequence were used in conjunction with a common test primer (C1, C2, K1, K2 or primers for the *tetRA* cassette) to test for both new-junction fragments.

Complementation studies. The *mchDEF* region was amplified from EcN genomic DNA using primers listed in Supplementary Table 5. A region of 4,148 bp was amplified to ensure all regulatory elements were included. The PCR fragment was cloned into plasmid pCR-XL-TOPO using Zero Blunt TOPO PCR Cloning Kit (Invitrogen) according to the manufacturer's protocol, subcloned into the multiple cloning site of low-copy plasmid pWSK29, and confirmed by Sanger sequencing (Eton Bioscience). The *mchDEF mcmI* and *mchDEF mchI* constructs were amplified from EcN genomic DNA using primers listed in Supplementary Table 5. PCR fragments of 4,321 bp (*mchDEFmcmI*), 4,000 bp (*mchDEF*) and 400 bp (*mchI*) were directly assembled with plasmid pWSK30 using the Gibson assembly method.

***In vitro* growth assays.** The role of EcN microcins against cEc, AIEC and STm was tested by an *in vitro* competitive growth assay in iron-rich and iron-limiting conditions. Strains were grown in nutrient broth supplemented with 0.2 mM 2,2'-dipyridyl (Sigma) aerobically at 37 °C overnight. Approximately 5 × 10³ CFU ml⁻¹ from an overnight culture were inoculated into 0.1 ml of tissue-culture medium (DMEM/F12 with 10% fetal bovine serum (FBS); Invitrogen), as previously described^{23,24}. When indicated, medium was prepared with 1 μM ferric iron citrate (Sigma). Wild-type EcN or EcN mutants were inoculated in competition with wild-type STm, STm mutants, cEc, or AIEC. CFUs of each strain were enumerated by plating serial dilutions at 0, 5, 8, 11 and (for some assays) 16 h after inoculation.

Mice. The Institutional Animal Care and Use Committee at the University of California, Irvine approved all mouse experiments. Eight-to-ten-week-old male and female C57BL/6 *Slc11a1*⁺ (also known as *Nramp1*) mice were bred and housed in a vivarium. For some experiments, mice were purchased from Jackson laboratory and, upon arrival, allowed to acclimate to the new environment for at least one week before the start of an experiment. All mice were housed under specific pathogen-free conditions. Mice were fed an irradiated 2920X Teklad diet (Envigo). For one experimental procedure, mice were fed an irradiated iron-rich diet (1% diet FeSO₄ 2020, Teklad TD 160234; Envigo). Culturing faeces on MacConkey agar indicated that culturable enterobacteria were absent from our mice. Five to ten mice were used per experimental group, as indicated by the number of symbols in each figure panel. Group sizes were based on previous studies that detected a tenfold difference with statistical significance. For one experiment, ten-week-old Swiss Webster germ-free mice were used. Swiss Webster germ-free mice were maintained

in sterile isolators and were fed a double-irradiated diet (Purina 5066, Charles River Rodent 18% Vac Pac). For this experimental procedure, three mice per experimental group were used (see figure panels). Mice were randomly grouped in cages of a maximum five animals per cage. Similar numbers of male and female mice were used in each experimental group. No blinding was performed, with the exception of histopathology.

Dextran sulphate sodium-induced colitis mouse model. Drinking water was replaced with either filter-sterilized water (mock treatment) or with a filter-sterilized solution of 4% (w/v) dextran sulphate sodium (DSS; relative molecular mass 36,000–50,000; MP Biomedicals) in water as indicated. For DSS-treated mice, 3 days before the end of the experiment, drinking water was switched for 24 h to filter-sterilized water. Drinking water was then replaced with either filter-sterilized water (mock treatment) or with a filter-sterilized solution of 2% (w/v) DSS. At 4 days after the start of DSS treatment, animals were orally inoculated with a 1:1 ratio of 5 × 10⁸ CFU of wild-type EcN and mutants in competition with cEc or AIEC strains (resuspended in 0.1 ml LB broth). Faecal material was collected each day after bacterial administration, resuspended in sterile PBS and the bacterial load was determined by plating serial tenfold dilutions on selective agar plates. Animals were euthanized 5 days after inoculation. Caecal tissue was collected and then flash-frozen in liquid nitrogen and stored at –80 °C for later isolation of mRNA and protein. The caecal tip was fixed in 10% formalin for histopathology. Faecal material and caecal content were collected in sterile PBS and the bacterial load for the *E. coli* strains was determined by plating serial tenfold dilutions on selective agar plates. At some time points, a mouse did not produce faecal material, hence an *n* value is displayed for a given time point that is less than the experiment's indicated *n* value for that group. To differentiate EcN from other *E. coli* strains in biological samples, the strains were differentially marked with the low-copy number pACY-omega (Cm) or pHP45omega (Carb) (Supplementary Table 4), and markers were swapped in some experiments. Plasmids were stably maintained throughout the infection and similar colonies were counted in other differential media such as on MacConkey agar (not shown). When noted, the ratio of two strains in the faeces and caecal content was calculated by dividing the output ratio (EcN CFU/cEc or AIEC CFU) by the input ratio (EcN CFU/cEc or AIEC CFU). The competitive index was calculated for isogenic strains by dividing the output ratio (EcN wild-type CFU:mutant CFU) by the input ratio (EcN wild-type CFU:mutant CFU).

S. Typhimurium-induced colitis. C57BL/6 *Nramp1*⁺ mice were treated with streptomycin (100 μl of a 200 mg ml⁻¹ solution in sterile water) one day before infection. The following day, mice were orally inoculated with 1 × 10⁹ CFU of STm (resuspended in 0.1 ml LB broth) or with a 1:1 ratio of 5 × 10⁸ CFU each of STm and wild-type EcN, or EcN *mchDEF* strain for competitive colonization experiments. For therapeutic administration of EcN after STm infection, C57BL/6 *Nramp1*⁺ mice were treated with streptomycin (100 μl of a 200 mg ml⁻¹ solution in sterile water) one day before STm infection. The following day, mice were orally inoculated with 1 × 10⁹ CFU of STm. At days 2 and 5 after STm infection, mice were either mock-treated or treated with 1 × 10⁹ CFU of wild-type EcN or EcN *mchDEF*. Faecal material was collected each day after bacterial administration and resuspended in sterile PBS. Bacterial load for STm and EcN strains was then determined by plating serial tenfold dilutions on selective agar plates. At some time points, a mouse did not produce faecal material, hence an *n* value is displayed for a given time point that is less than the experiment's indicated *n* value for that group. At day 7 following infection, mice were euthanized and a portion of the caecum was flash-frozen in liquid nitrogen then stored at –80 °C for later isolation of mRNA and protein. The caecal tip was fixed in 10% formalin for histopathology. Bacteria in the faecal content were counted by plating serial tenfold dilutions on LB agar plates containing the appropriate antibiotics. To selectively identify STm from EcN strain in the faecal content, strains were differentially marked with the low-copy number pACYomega (Cm) or pHP45omega (Carb) plasmids. Plasmids were stably maintained throughout the infection and similar colonies were counted in other differential media such as MacConkey agar (not shown). When noted, the ratio of two strains was calculated by dividing the output ratio (EcN CFU/STm CFU) by the input ratio (EcN CFU/STm CFU).

Extraction of bacterial DNA from faecal samples. Faecal samples were collected before DSS treatment (day –4), at day 4 after DSS treatment (day 0; before inoculation of bacteria) and at day 5 after administration of commensal *E. coli* in competition with wild-type EcN or EcN *mchDEF*. Faeces were snap-frozen in liquid nitrogen, then DNA was later extracted using the QIAamp DNA Stool Kit (Qiagen) according to the manufacturer's instructions with modifications as previously described¹⁸. Because DSS interferes with PCR, we used the following procedure (developed by W. Zhu and S. Winter) to eliminate DSS from samples. Samples were eluted in 200 μl of ultra-pure water and combined with 80 mg of KCl followed by vortexing for 2 min. The mixture was then incubated on ice for 30 min and centrifuged at 16,800 g at 4 °C for 30 min. The supernatant was then transferred

to a new tube and 1/10 volume of 3 M NaAc (pH 5.2) was added. We then added 2.5 ml of pure ethanol to the mixture, vortexed and centrifuged at 16,800g at 4 °C for 30 min. The supernatant was removed and 1 ml of 70% ethanol was added to the pellet, vortexed and centrifuged at 16,800g for 5 min at room temperature. This wash step was repeated 3–4 times. The pellet was then air-dried at 55 °C and resuspended in 30 µl of ultra-pure water.

Analysis of microbiota. DNA extracted from faecal samples was amplified by a two-step PCR enrichment of 16S rDNA (V4 region) with primers 515F and 806R modified by addition of barcodes for multiplexing, then sequenced on an Illumina MiSeq system (UC Davis HMSB Facility). Sequences were processed and analysed by employing the QIIME pipeline v1.9.1 (ref. 28) with default settings, except as noted. In brief, paired-end sequences were joined, quality-filtered, reverse-complemented and chimera-filtered (usearch61 option; RDP gold database); operational taxonomic units (OTUs) were picked (usearch61 option; enable_rev_strand_match True) at 97% similarity; taxonomy was assigned (confidence 0.8) with the RDP classifier. Greengenes database v13_8 was used in the open-reference OTU picking workflow.

Bacterial RNA extraction. A bacterial RNA mini-kit (Bio-Rad Aurum Total) was used to extract RNA from bacterial cultures. EcN strains were grown in nutrient broth supplemented with 0.2 mM 2,2'-dipyridyl aerobically at 37 °C overnight. Approximately 10^4 CFU ml⁻¹ from an overnight culture were inoculated into 5 ml of tissue culture medium (DMEM/F12 plus 10% FBS, Invitrogen), as previously described^{23,24}. When indicated, iron citrate was added to the medium at a final concentration of 1 µM. At 7 h post-inoculation, 2×10^9 CFU were used to extract RNA. An additional DNase treatment (Ambion) was done before the generation of cDNA with reverse transcription reagents (Roche).

Quantitative real-time PCR. For analysis of gene expression by quantitative real-time PCR, total RNA was extracted from caecal tissue with TRI Reagent (Molecular Research Center). RNA from DSS-treated mice was further purified using the Dynabeads mRNA DIRECT Purification Kit (Life Technologies) according to manufacturer recommendations. Reverse-transcription reagents (Roche) were employed to generate cDNA from all RNA samples. Real-time PCR was performed using SYBR Green (Roche) and the Roche Lightcycler 480 Instrument II system (Roche). Data were analysed using the comparative $-2^{\Delta\Delta C_t}$ method. Target gene transcription of each tissue sample was normalized to the respective levels of *Actb* mRNA (β -actin). For qPCR analysis of bacterial transcripts, transcription of

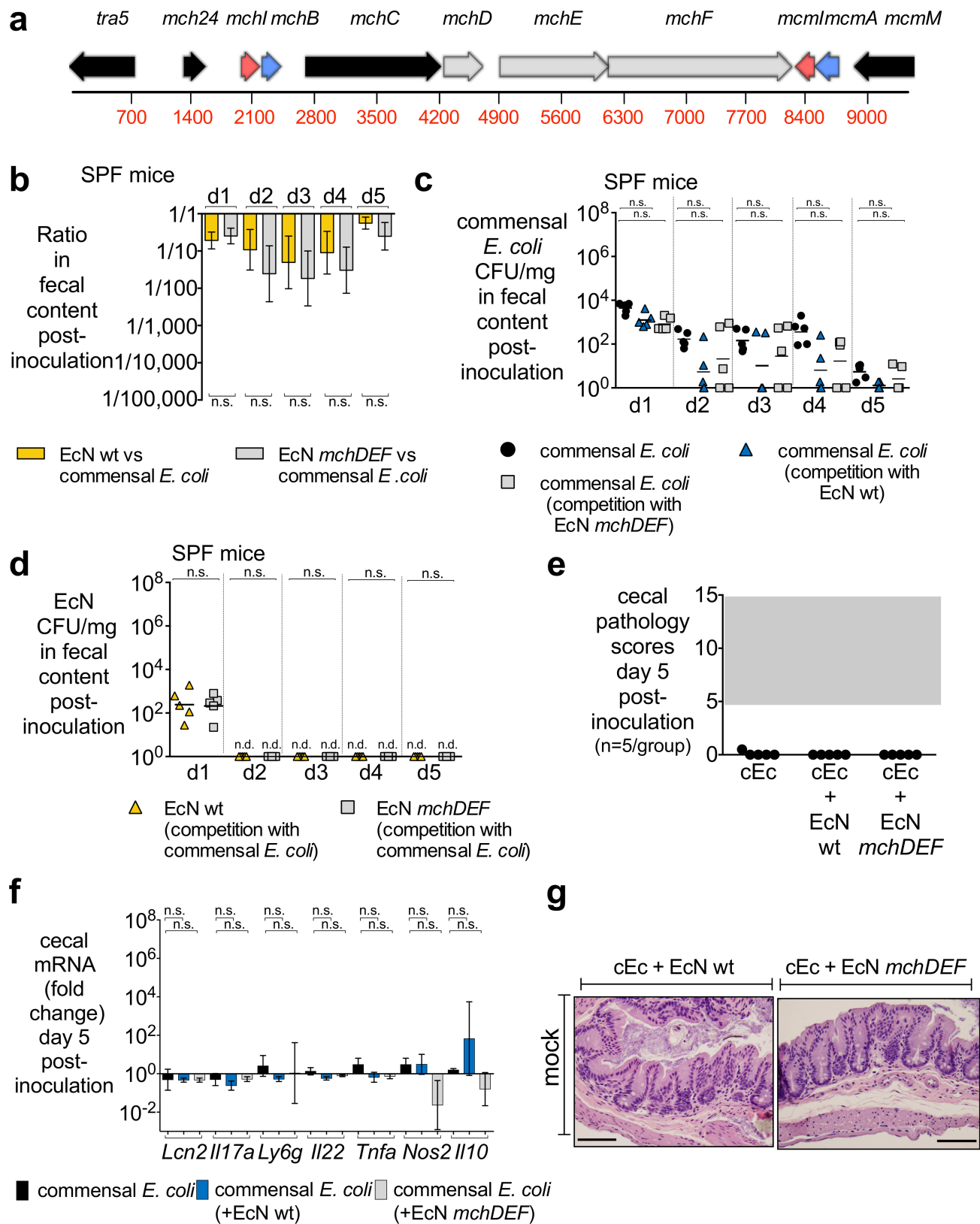
mcmA and *mchB* was normalized to bacterial *gapA* mRNA levels. Data represent at least three independent experiments. DNA contamination was less than 1% for all bacterial amplicons, as determined by separate mock reactions lacking reverse transcriptase. All primers used are listed in Supplementary Table 6.

Histopathology analysis. Tissue samples were fixed in 10% formalin, processed according to standard procedures for paraffin embedding, sectioned at 5 mm, and stained with haematoxylin and eosin. The pathology score of caecal samples was determined by blinded examination of caecal sections from a board-certified pathologist using previously published methods¹⁸. Each section was evaluated for the presence of neutrophils, mononuclear infiltrate, submucosal oedema, surface erosions, inflammatory exudates, and cryptitis. Inflammatory changes were scored from 0 to 4 according to the following scale: 0, none; 1, low; 2, moderate; 3, high; 4, extreme. The inflammation score was calculated by adding up all scores obtained for each parameter and interpreted as follows: 0–2, within normal limit; 3–5, mild; 6–8, moderate; ≥ 8 , severe.

Statistical analysis of data. The experiments were not randomized. No statistical methods were used to predetermine sample size. Prism 6 and 7 software (GraphPad) was used for statistical analysis. Bacterial growth curves were analysed by unpaired Student's *t*-test. To compare bacterial CFUs in faeces and caecal content, mRNA expression, and eubacterial taxa present at greater than 0.5% relative abundance in at least one sample, we applied a non-parametric Mann–Whitney–Wilcoxon test. We chose this test because it can be applied to data with normal or unknown distribution. *P*-values for all statistical comparisons presented in the main text, figures and Extended Data are available in Supplementary Table 7.

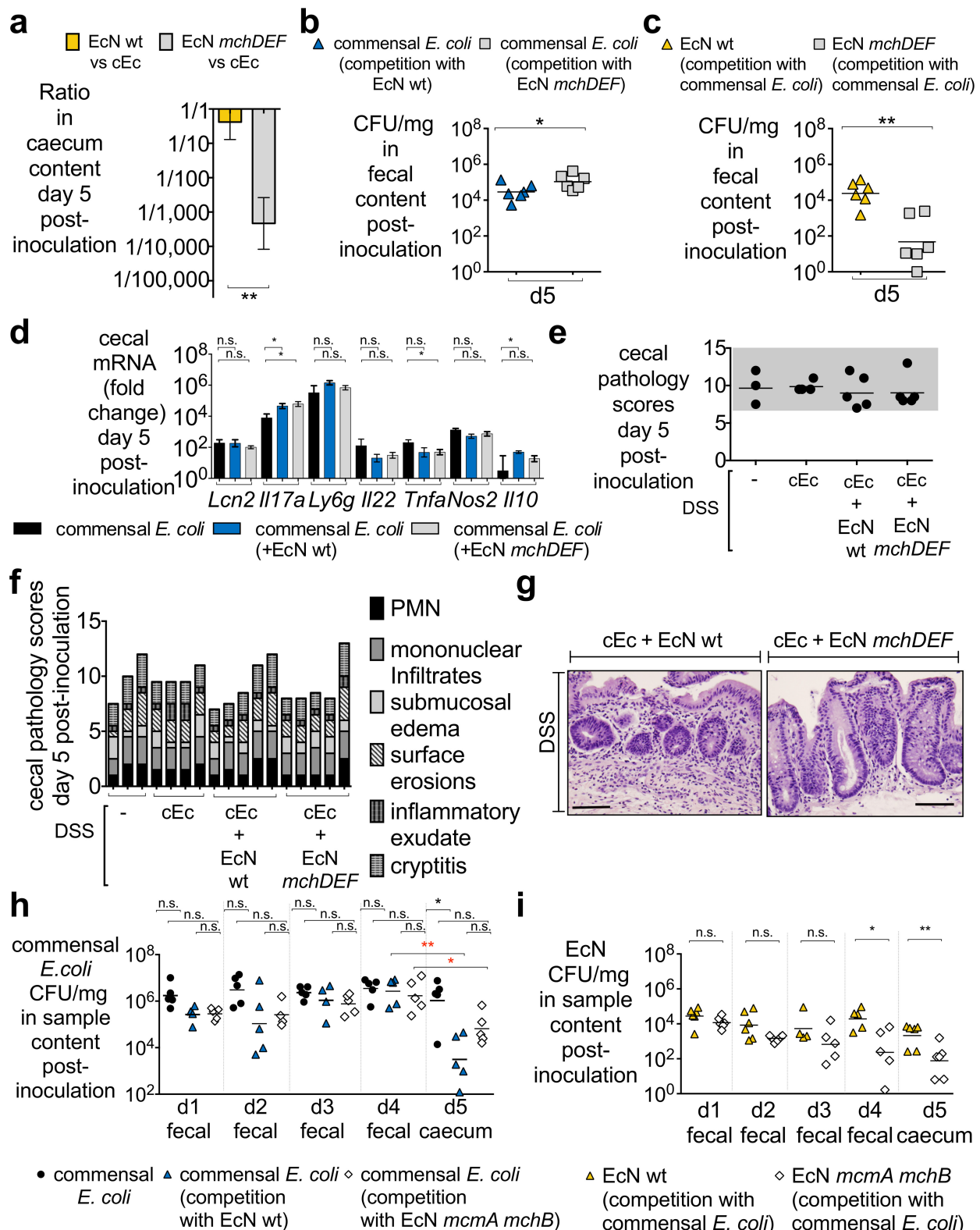
Data availability. The data that support the findings of this study are available from the corresponding author upon request. Faecal microbiota 16S rRNA gene sequencing data were deposited in the European Nucleotide Archive under accession PRJEB15700.

26. Eaves-Pyles, T. *et al.* *Escherichia coli* isolated from a Crohn's disease patient adheres, invades, and induces inflammatory responses in polarized intestinal epithelial cells. *Int. J. Med. Microbiol.* **298**, 397–409 (2008).
27. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
28. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).



Extended Data Figure 1 | Wild-type EcN and microcin mutant (EcN *mchDEF*) gut colonization in competition with commensal *E. coli* in the absence of intestinal inflammation. **a**, Illustration of the microcin gene cluster in EcN. **b–g**, Intra-gastric inoculation of SPF mice with cEc alone or in competition with wild-type EcN or EcN *mchDEF* at a 1:1 ratio. **b**, Ratio (EcN over cEc) in faecal content at days 1–5 after inoculation ($n = 5$ per group). **c**, **d**, CFU mg^{-1} of (c) cEc or (d) wild-type EcN or EcN *mchDEF* in faecal content at days 1–5 after inoculation, when cEc was administered

alone or in competition as indicated. **e**, **f**, Caecal histopathology scores (e) or gene expression (f) (data are expressed as fold change over mock-treated mice) at day 5 after inoculation for panels **b–d** ($n = 5$ per group). **g**, Haematoxylin and eosin-stained sections from representative animals at day 5 after inoculation. Scale bar, 100 μm . Each individual symbol represents one mouse (c–e). Bars represent the geometric mean \pm s.e.m. (b, f), geometric mean (c, d), or mean (e). n.s., not significant.

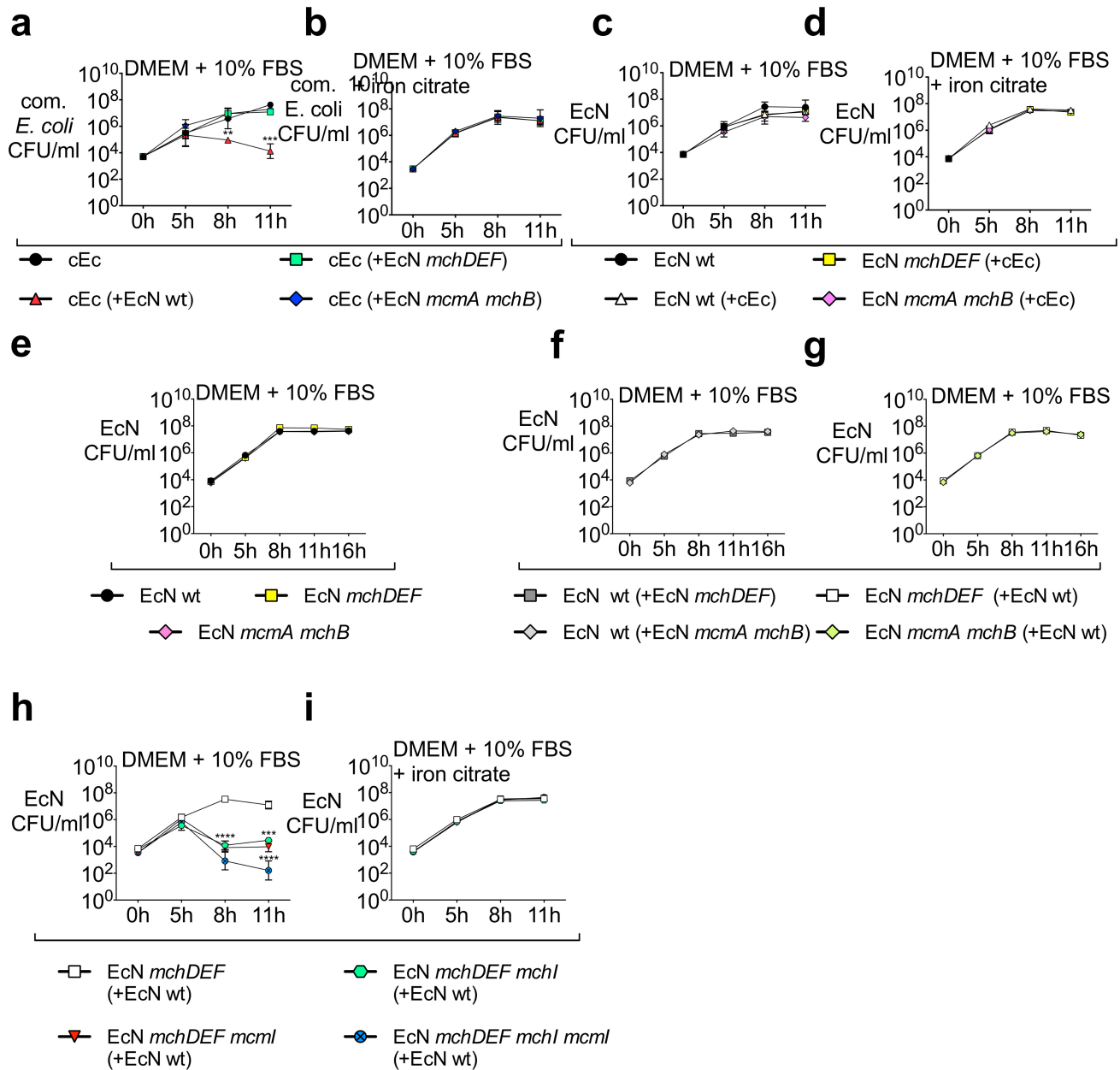


Extended Data Figure 2 | Colonization and histopathology of wild-type EcN and microcin mutants (*mchDEF* and *mcmA mchB*) in competition with commensal *E. coli*, in mice with DSS-mediated colitis.

a–i, Experimental design as in Fig. 2a with SPF mice. **a**, Ratio (EcN over cEc) in caecal content at day 5 after intragastric inoculation ($n = 6$ per group). **b**, **c**, CFU mg^{-1} of (b) cEc or (c) wild-type EcN or EcN *mchDEF* in faecal content at day 5 after inoculation when in competition as indicated ($n = 6$ per group). **d**, **e**, Caecal (d) gene expression ($n = 5$ per group; data are presented as fold change over mock-treated mice) or (e) histopathology scores at day 5 after inoculation from mice shown in

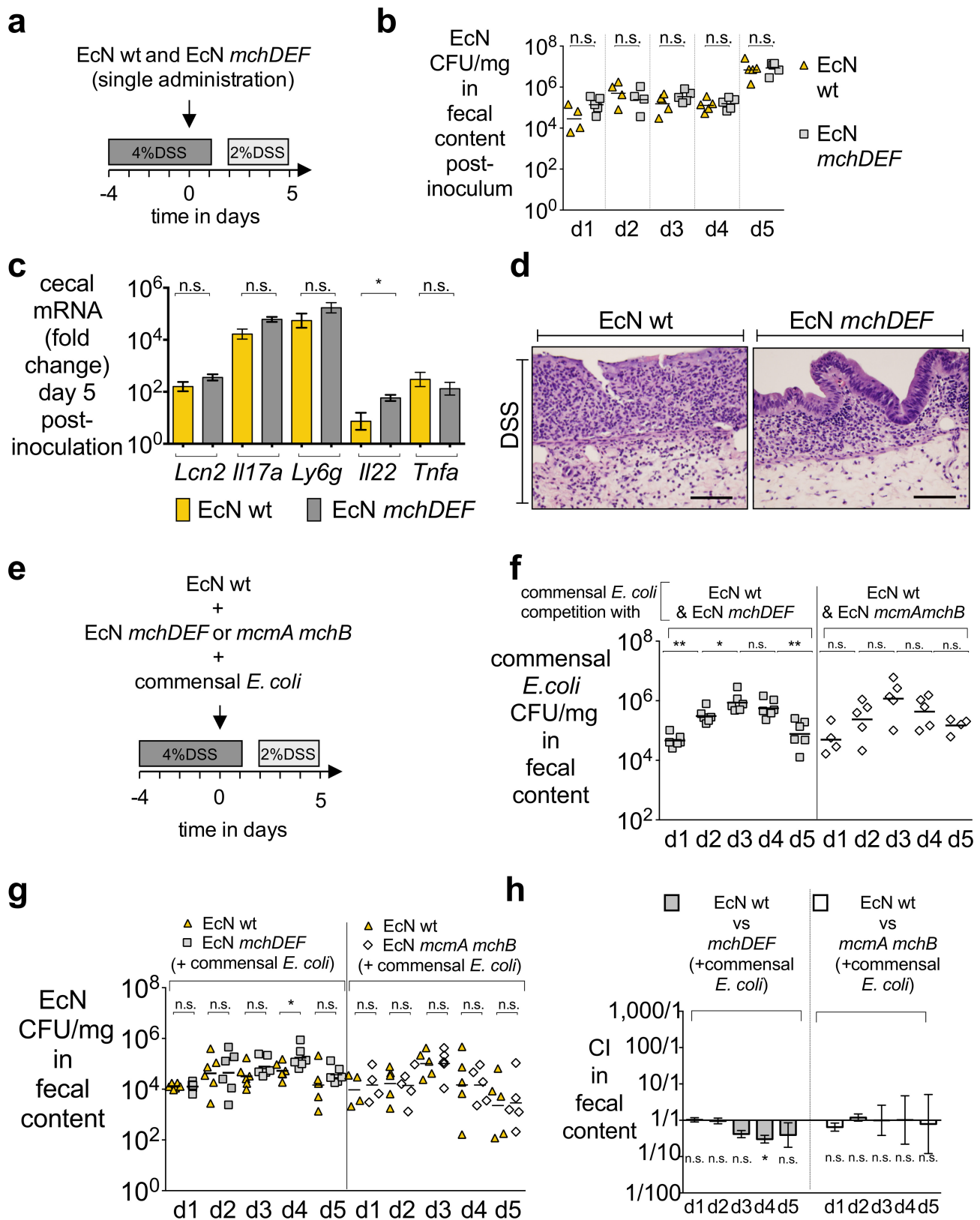
panels **b**, **c** and Fig. 2b–d (DSS only, $n = 3$; all others, $n = 5$).

Scale bar, 100 μm . **f**, Detailed histopathology scoring of mice in **e**. **g**, Haematoxylin and eosin-stained sections from representative animals at day 5 following inoculation. **h**, **i**, CFU mg^{-1} of (h) cEc or (i) wild-type EcN or EcN *mcmA mchB* in sample content at days 1–5 following intragastric inoculation when cEc was administered alone or in competition as indicated ($n = 5$ per group). Each individual symbol represents one mouse (**b**, **c**, **e**, **h**, **i**). Bars represent the geometric mean \pm s.e.m. (**a**, **d**), geometric mean (**b**, **c**, **h**, **i**) or mean (**e**). * $P < 0.05$, ** $P < 0.01$; n.s., not significant.



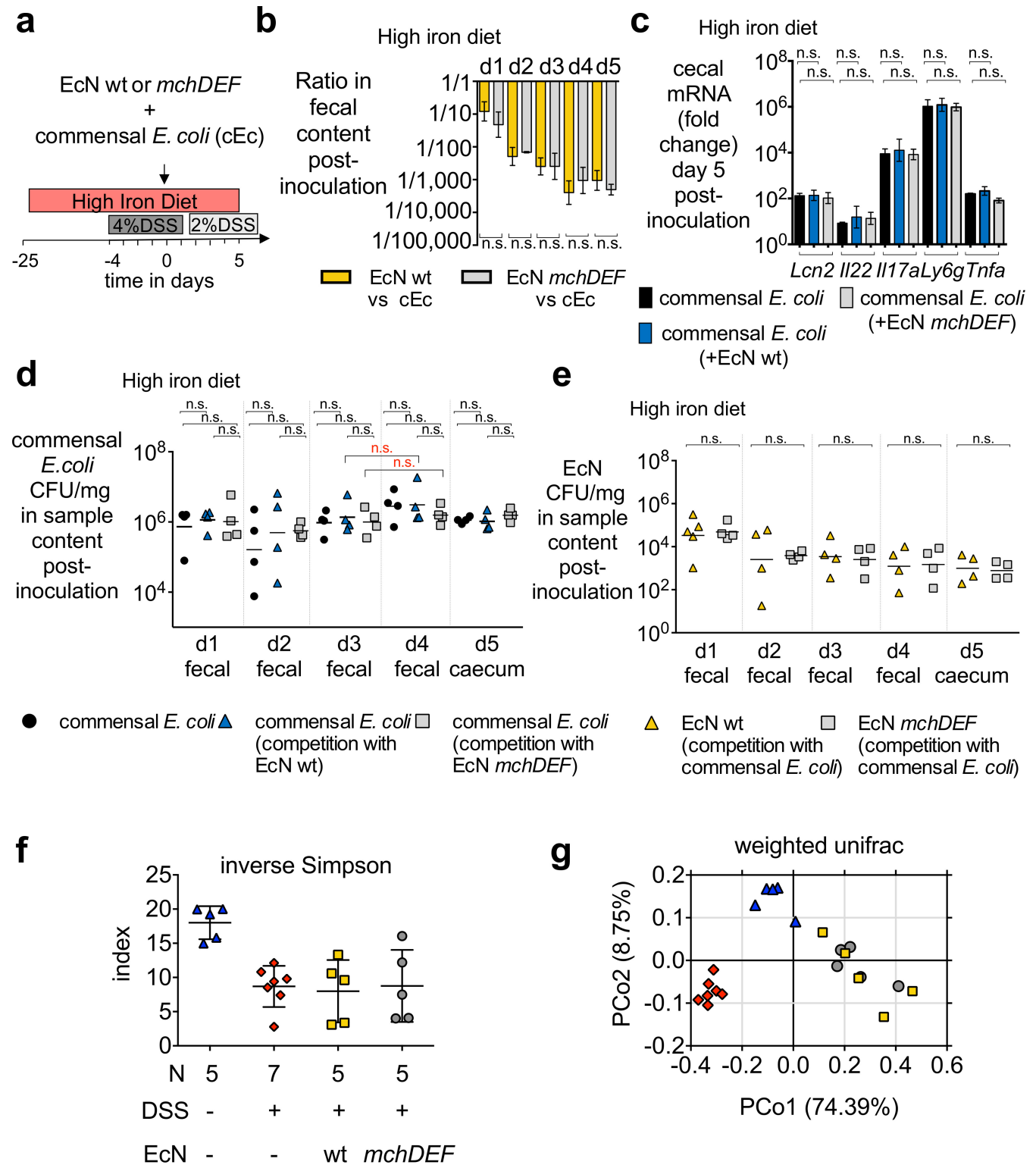
Extended Data Figure 3 | In vitro growth curves of EcN and microcin mutants when grown alone or in competition with EcN mutants or commensal *E. coli*. a–i, Strains were grown overnight in nutrient broth supplemented with 0.2 mM 2,2'-dipyridyl. Growth assays were performed in iron-limiting conditions (DMEM/F12 supplemented with 10% FBS) or in medium supplemented with 1 μ M iron citrate. Time points at 0, 5, 8 and 11 h after inoculation were collected. a, b, cEc CFU ml⁻¹ when grown alone or in competition with wild-type or mutant EcN in iron-limiting conditions (a) or in medium supplemented with 1 μ M iron citrate (b). c, d, CFU ml⁻¹ of wild-type or mutant EcN when grown alone or in

competition with cEc in iron-limiting conditions (c) or in medium supplemented with 1 μ M iron citrate (d). e–g, Under iron-limiting conditions, CFU ml⁻¹ of wild-type or mutant EcN when grown alone (e), of wild-type EcN (f) or EcN *mchDEF* or EcN *mcmA mchB* (g) when grown competitively as indicated. h, i, CFU ml⁻¹ of the indicated EcN mutants (immunity gene(s) and/or *mchDEF*) grown in competition with wild-type EcN in (h) iron-limiting conditions or in (i) medium supplemented with 1 μ M iron citrate. Symbols represent the geometric mean of three independent experiments \pm s.e.m. ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.



Extended Data Figure 4 | Gut colonization of wild-type EcN and microcin mutants in the DSS-colitis model. **a**, Experimental design for **b–d** with SPF mice. **b**, CFU mg^{-1} of wild-type EcN or EcN *mchDEF* in faecal content at days 1–5 following intragastric inoculation ($n = 5$ per group). **c**, **d**, Caecal gene expression (**c**; $n = 5$ per group; data are expressed as fold change over mock-treated mice) or haematoxylin and eosin-stained sections (**d**; representative) from panel **b** mice at day 5 following inoculation. Scale bar, 100 μm . **e**, Triple co-administration

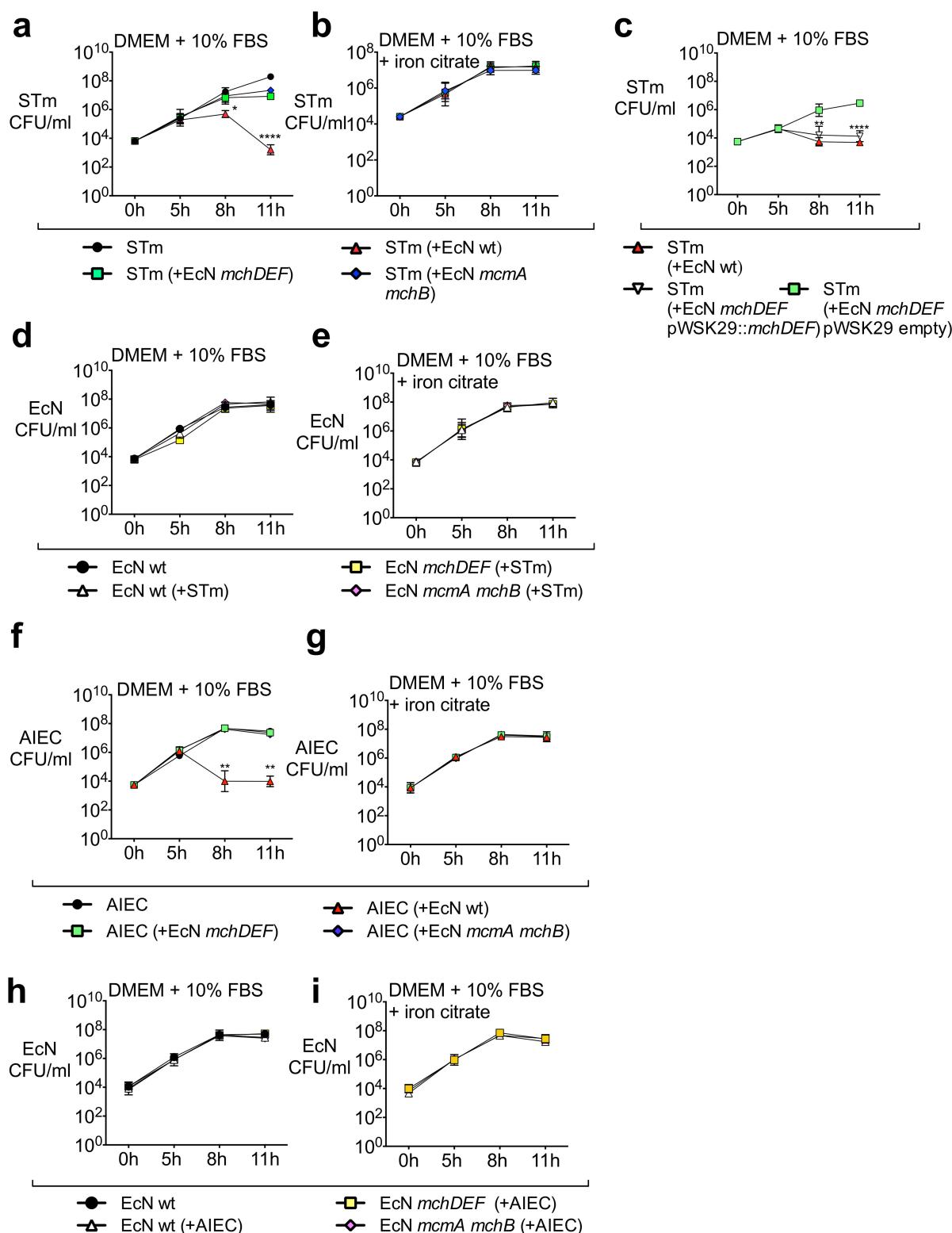
design for **f–h** with SPF mice. **f**, **g**, CFU mg^{-1} of cEc (**f**) or indicated EcN strain (**g**) in faecal content at days 1–5 following intragastric inoculation with cEc, wild-type EcN and the indicated EcN mutant ($n = 5$ per group). **h**, Competitive index (CI; EcN wild-type over mutant) for EcN data presented in panel **g** ($n = 5$ per group). **b**, **f**, **g**, Each individual symbol represents one mouse. Bars represent the geometric mean (**b**, **f**, **g**) or the geometric mean \pm s.e.m. (**c**, **h**). * $P < 0.05$, ** $P < 0.01$; n.s., not significant.



Extended Data Figure 5 | Impact of microcins on the microbiota and impact of a high iron diet on microcin-mediated competition.

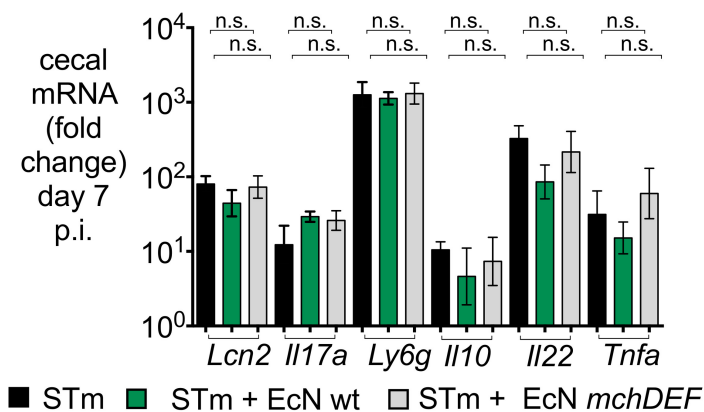
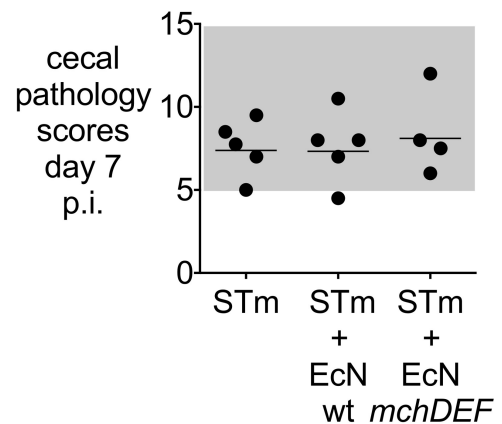
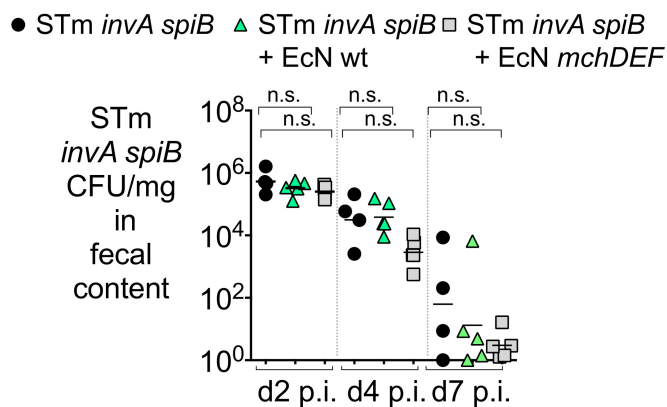
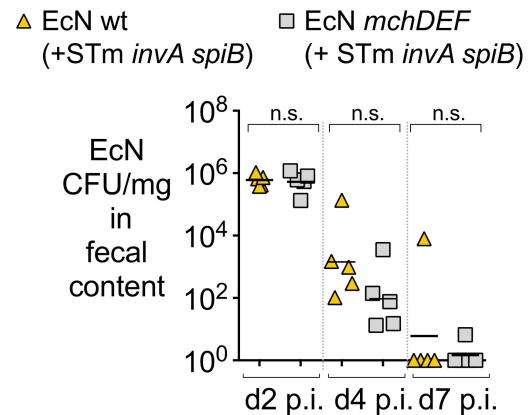
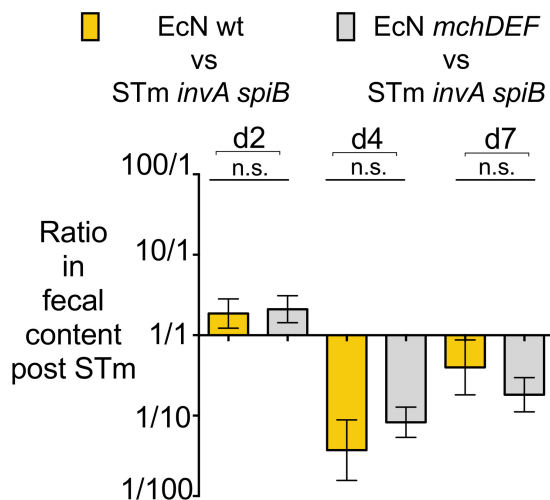
a, High-iron diet design for **b–e** with SPF mice ($n = 5$ per group) **b**, Ratio (EcN over cEc) in faecal content at days 1–5 after intragastric inoculation. **c**, Caecal gene expression at day 5 following inoculation ($n = 3–4$ per group). **d**, **e**, CFU mg^{-1} of cEc (**d**) or wild-type EcN or EcN *mchDEF* (**e**) in indicated samples at days 1–5 after inoculation when cEc was administered alone or in competition as indicated ($n = 5$ per group). **f**, **g**, See Fig. 2a for experimental design with SPF mice. 16S ribosomal rRNA gene sequence analysis (V4 region) of faecal DNA obtained from

mice prior to DSS administration (day –4), after DSS administration (Day 0), and day 5 after intragastric inoculation of cEc with either wild-type EcN or EcN *mchDEF* at a 1:1 ratio. **f**, Eubacterial alpha diversity (inverse Simpson index); Shannon index yielded similar results. **g**, Principal coordinates (PCo) analysis plot (PCo1 versus PCo2) of eubacterial beta diversity (weighted UniFrac); symbols as in panel **f**. Each individual symbol represents one mouse (**d**, **e**, **f**, **g**). Bars represent the geometric mean \pm s.e.m. (**b**, **c**), geometric mean (**d**, **e**) or mean \pm s.d. (**f**). n.s., not significant. Black symbols represent comparisons within same day; red symbols represent comparisons between days.



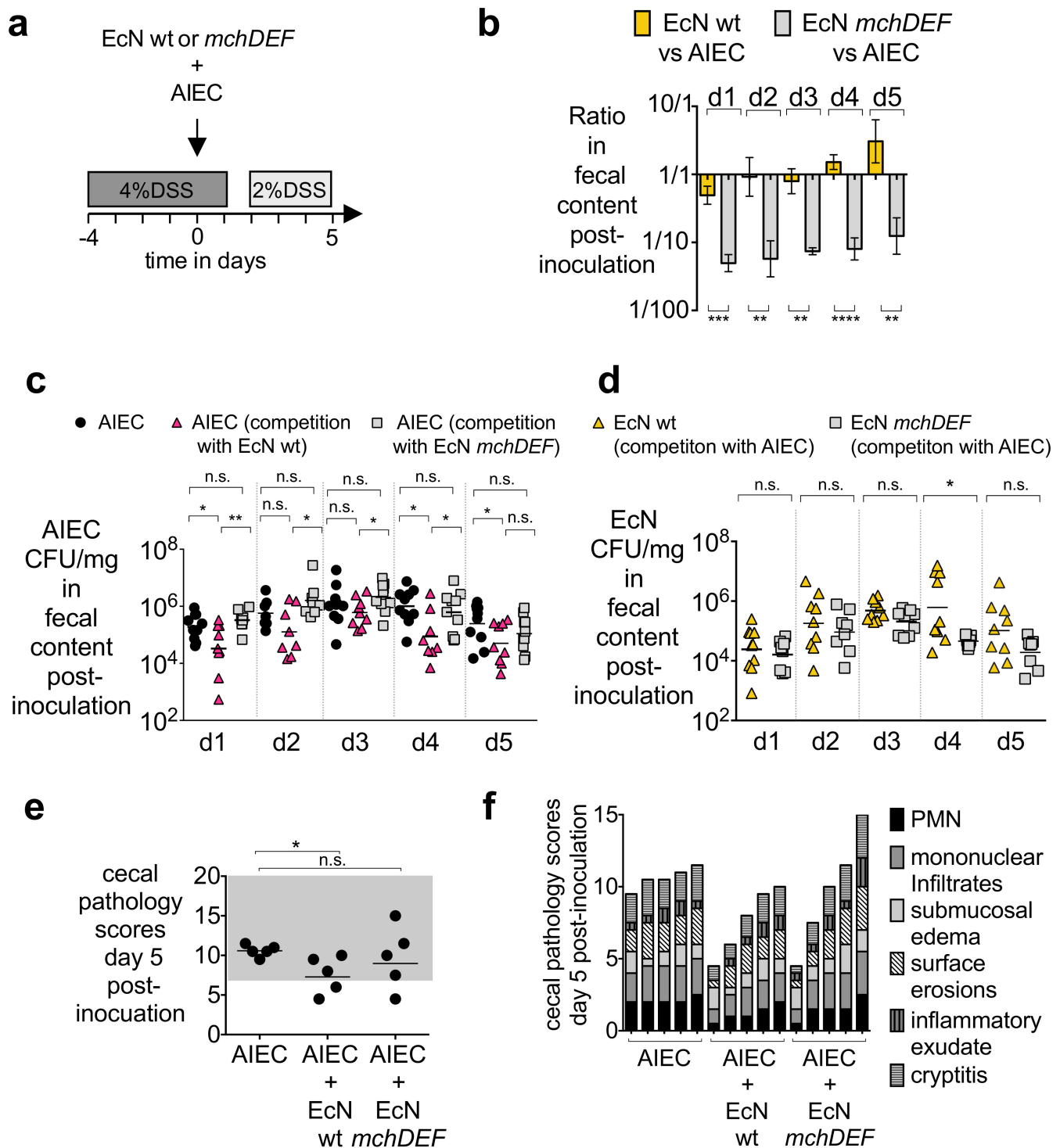
Extended Data Figure 6 | In vitro activity of EcN microcins against *S. Typhimurium* and AIEC. a–i, Strains were grown overnight in nutrient broth supplemented with 0.05M 2,2'-dipyridyl. Growth assays were performed in iron-limiting conditions (DMEM/F12 supplemented with 10% FBS) or in medium supplemented with 1 μ M iron citrate. Time points at 0, 5, 8 and 11 h after inoculation were collected. a, b, STm CFU ml⁻¹ when grown alone or in competition with wild-type or mutant EcN in iron-limiting conditions (a) or in medium supplemented with 1 μ M iron citrate (b). c, STm CFU ml⁻¹ in iron-limiting conditions when in competition with wild-type EcN or an EcN *mchDEF* strain harbouring either pWSK29::*mchDEF* or the empty-vector control. d, e, CFU ml⁻¹ of

wild-type or mutant EcN when grown alone or in competition with STm in (d) iron-limiting conditions or in (e) medium supplemented with 1 μ M iron citrate. f, g, AIEC CFU ml⁻¹ when grown alone or in competition with wild-type or mutant EcN in iron-limiting conditions (f) or in medium supplemented with 1 μ M iron citrate (g). h, i, CFU ml⁻¹ of wild-type or mutant EcN when grown alone or in competition with AIEC in iron-limiting conditions (h) or in medium supplemented with 1 μ M iron citrate (i). Symbols represent the geometric mean (three independent experiments) \pm s.e.m. * P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001.

a**b****c****d****e**

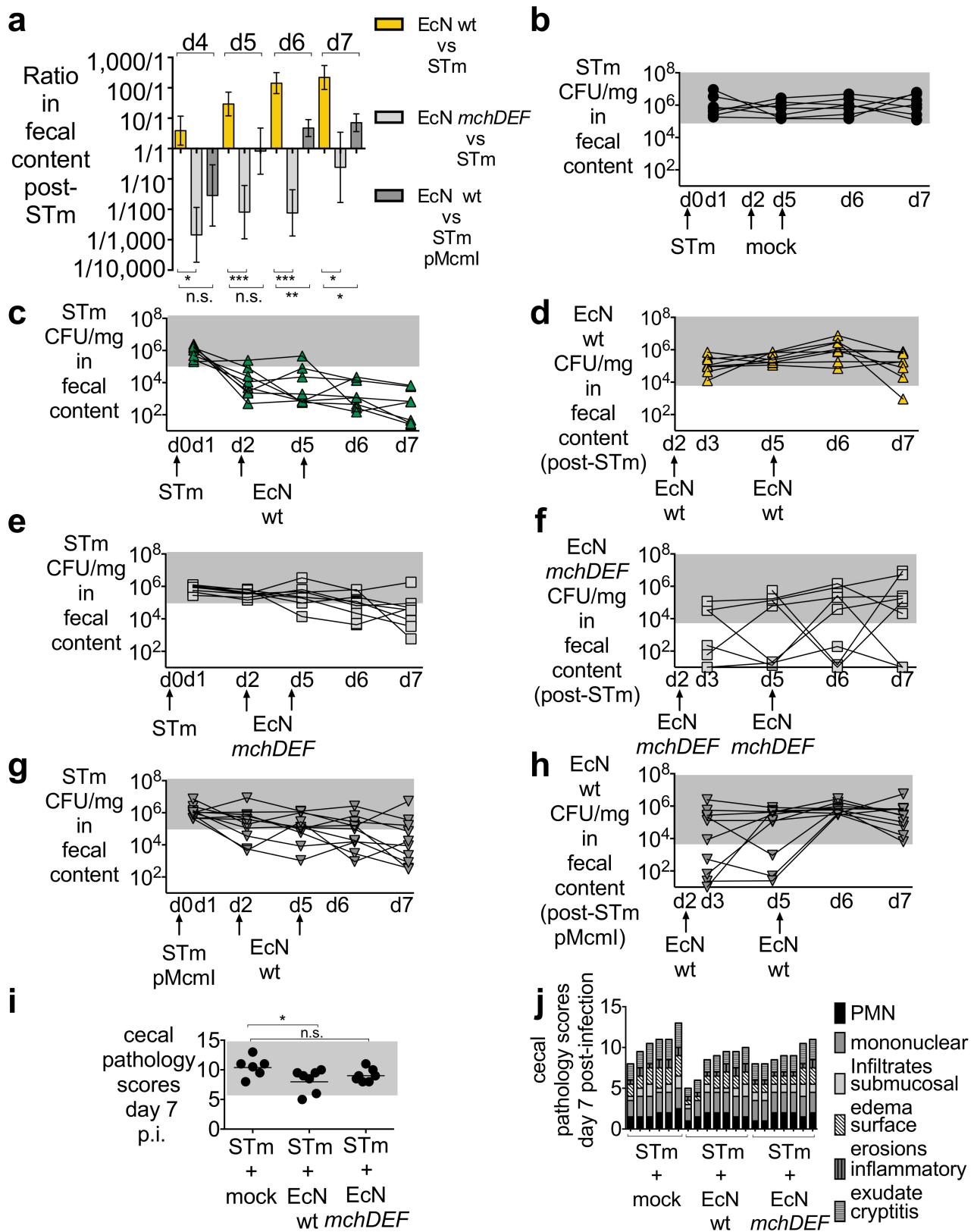
Extended Data Figure 7 | Co-administration of *S. Typhimurium* with wild-type EcN or EcN *mchDEF*. a–e, See Fig. 3d for co-administration design with SPF mice. a, b, Intragastric inoculation with wild-type STm. Caecal (a) gene expression (STm only, $n = 7$; all others, $n = 9$); data are expressed as fold change over mock-treated mice or (b) histopathology (STm only and STm + wild-type EcN, $n = 5$; STm + EcN *mchDEF*, $n = 4$) from mice shown in Fig. 3f, g at Day 7 following infection. c–e, Intragastric inoculation with STm *invA spiB*. c, d, CFU mg^{-1} of (c)

STm *invA spiB* ($n = 4$) or (d) wild-type EcN or EcN *mchDEF* in faecal content at designated time points after infection when STm *invA spiB* was administered alone or in competition as indicated ($n = 5$ per group). e, Ratio of wild-type EcN or EcN *mchDEF* over STm *invA spiB* in faecal content at designated time points after infection ($n = 5$ per group). Each individual symbol represents one mouse (b–d). Bars represent the geometric mean \pm s.e.m. (a, e), mean (b), or geometric mean (c, d). n.s., not significant.



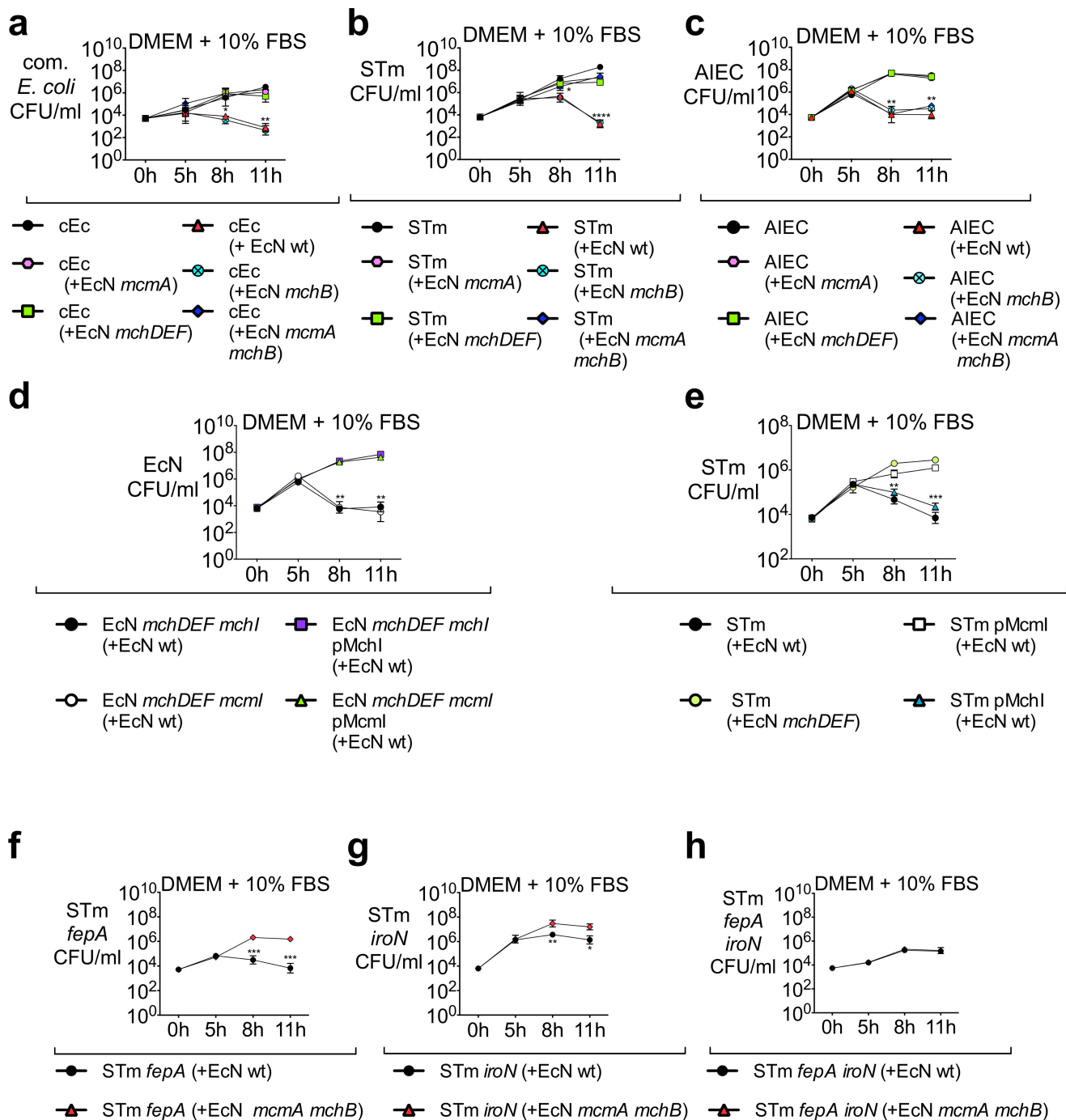
Extended Data Figure 8 | Gut colonization of AIEC in DSS-treated mice when competing with wild-type EcN or EcN *mchDEF*. **a**, Experimental design for **b–f** with SPF mice. **b**, Ratio of wild-type EcN or EcN *mchDEF* over AIEC in faecal content at days 1–5 following intragastric inoculation ($n=9$ per group). **c**, **d**, CFU mg^{-1} of (c) AIEC or (d) wild-type EcN or EcN *mchDEF* in faecal content at days 1–5 following inoculation when

AIEC was administered alone or in competition as indicated ($n=9$ per group). **e**, Caecal histopathology scores at day 5 after inoculation for panels **c**, **d** ($n=5$ per group). **f**, Detailed histopathology scoring of panel **e** mice. Each individual symbol represents one mouse (**c–e**). Bars represent the geometric mean \pm s.e.m. (**b**), geometric mean (**c**, **d**) or mean (**e**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$; n.s., not significant.



Extended Data Figure 9 | Therapeutic administration of wild-type EcN, EcN *mchDEF* or mock during *S. Typhimurium* infection. a–j. See Fig. 4a for therapeutic design with SPF mice. a–f, Intragastric inoculation with wild-type STm. a, Ratio of wild-type EcN or EcN *mchDEF* over STm in faecal content on days 4–7 after infection with STm ($n=8$ per group). b, c, e, STm CFU mg^{-1} at designated time points after infection in faecal content of mice therapeutically treated with mock (b; $n=7$), wild-type EcN (c; $n=8$) or EcN *mchDEF* (e; $n=8$). d, f, CFU mg^{-1} of wild-type EcN (d) or EcN *mchDEF* (f) in faecal content at designated time points after

STm infection. g, h, Intragastric inoculation with STm pMcmI ($n=10$ per group). CFU mg^{-1} of STm pMcmI (g) or wild-type EcN (h) in faecal content at designated time points following STm pMcmI infection. Grey box represents average STm CFU mg^{-1} in mock-treated mice (panel b). i, Cecal histopathology scores for panels b–f (STm alone, $n=6$; all others, $n=7$). j, Detailed histopathology scoring of mice in i. Each individual symbol represents one mouse (b–h). Bars represent the geometric mean \pm s.e.m. (a) or mean (i). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; n.s., not significant.



Extended Data Figure 10 | In vitro growth curves of microcin M (*mcmA*) and microcin H47 (*mchB*) EcN mutants in competition with commensal *E. coli*, *S. Typhimurium*, or AIEC. a–h, Strains were grown overnight in Nutrient Broth supplemented with 0.2 mM 2,2'-dipyridyl. Growth assays were performed in iron-limiting conditions (DMEM/F12 supplemented with 10% FBS) and time points at 0, 5, 8 and 11 h following inoculation were collected. a–c, CFU ml⁻¹ of commensal *E. coli* (a), STm (b) or AIEC (c) when grown alone or in competition with the indicated

EcN strain. d, CFU ml⁻¹ of complemented and uncomplemented EcN microcin immunity gene mutants when in competition with EcN wild-type. e, CFU ml⁻¹ of wild-type STm or STm harbouring pMchI or pMcmI when in competition with EcN wild-type. f–h, CFU ml⁻¹ of STm *fepA* (f), STm *iroN* (g) and STm *fepA iroN* (h) in competition with either wild-type EcN or EcN *mcmA mchB*. Symbols represent the geometric mean for three independent experiments ± s.e.m. **P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001.

Ad26/MVA therapeutic vaccination with TLR7 stimulation in SIV-infected rhesus monkeys

Erica N. Borducchi¹, Crystal Cabral¹, Kathryn E. Stephenson¹, Jinyan Liu¹, Peter Abbink¹, David Ng'ang'a¹, Joseph P. Nkolola¹, Amanda L. Brinkman¹, Lauren Peter¹, Benjamin C. Lee¹, Jessica Jimenez¹, David Jetton¹, Jade Mondesir¹, Shanell Mojta¹, Abishek Chandrashekar¹, Katherine Molloy¹, Galit Alter², Jeffrey M. Gerold³, Alison L. Hill³, Mark G. Lewis⁴, Maria G. Pau⁵, Hanneke Schuitemaker⁵, Joseph Hesselgesser⁶, Romas Geleziunas⁶, Jerome H. Kim⁷†, Merlin L. Robb⁷, Nelson L. Michael⁷ & Dan H. Barouch^{1,2}

The development of immunologic interventions that can target the viral reservoir in HIV-1-infected individuals is a major goal of HIV-1 research^{1,2}. However, little evidence exists that the viral reservoir can be sufficiently targeted to improve virologic control following discontinuation of antiretroviral therapy. Here we show that therapeutic vaccination with Ad26/MVA (recombinant adenovirus serotype 26 (Ad26) prime, modified vaccinia Ankara (MVA) boost)^{3,4} and stimulation of TLR7 (Toll-like receptor 7) improves virologic control and delays viral rebound following discontinuation of antiretroviral therapy in SIV-infected rhesus monkeys that began antiretroviral therapy during acute infection. Therapeutic vaccination with Ad26/MVA resulted in a marked increase in the magnitude and breadth of SIV-specific cellular immune responses in virologically suppressed, SIV-infected monkeys. TLR7 agonist administration led to innate immune stimulation and cellular immune activation. The combination of Ad26/MVA vaccination and TLR7 stimulation resulted in decreased levels of viral DNA in lymph nodes and peripheral blood, and improved virologic control and delayed viral rebound following discontinuation of antiretroviral therapy. The breadth of cellular immune responses correlated inversely with set point viral loads and correlated directly with time to viral rebound. These data demonstrate the potential of therapeutic vaccination combined with innate immune stimulation as a strategy aimed at a functional cure for HIV-1 infection.

The critical barrier to the discovery of a cure for HIV-1 is the viral reservoir in latently infected CD4⁺ T lymphocytes^{5–8}, which leads to viral rebound in the vast majority of HIV-1-infected individuals following discontinuation of antiretroviral therapy (ART)^{9,10}. Enhancing antiviral immune responses, potentially together with activation of the viral reservoir, might be able to eliminate these cells^{12,11}. However, it is currently unknown whether immunologic interventions can affect the viral reservoir *in vivo*. In particular, it is unclear if a therapeutic vaccine will be able to induce cellular immune responses with sufficient potency and breadth to control viral rebound following ART discontinuation¹². We therefore evaluated a strategy consisting of Ad26/MVA therapeutic vaccination^{3,4} and TLR7 agonist GS-986 administration in ART-suppressed, SIV-infected rhesus monkeys.

We infected 36 Indian origin rhesus monkeys (*Macaca mulatta*) with SIV_{mac251} (refs 4, 13) by a single intrarectal exposure and initiated daily subcutaneous administration of a pre-formulated ART cocktail (tenofovir disoproxil fumarate, emtricitabine, dolutegravir)¹⁴ on day 7 of acute infection. Animals had median plasma SIV RNA levels of 7.10 log copies per ml (range 6.04–7.88 log copies per ml) on the day of ART initiation (Fig. 1a). SIV RNA levels were controlled in the majority of animals by day 56 and in all animals by day 224

(Fig. 1a). SIV RNA levels were comparable among the different groups, and the animals that took longer to control the virus had higher starting plasma viral loads on day 7 ($P = 0.04$; data not shown).

Following 24 weeks of suppressive ART, groups of monkeys received the following interventions: (1) Ad26/MVA vaccines alone; (2) Ad26/MVA vaccines and TLR7 agonist GS-986; (3) TLR7 agonist GS-986 alone; or (4) sham ($n = 9$ animals per group). In groups 1–2, animals were vaccinated by the intramuscular route with 3×10^{10} viral particles of Ad26 vectors^{4,15} expressing SIV_{smE543} gag-pol-env at weeks 24 and 36, and were boosted with 10^8 plaque-forming units of MVA vectors⁴ expressing SIV_{smE543} gag-pol-env at weeks 48 and 60. In groups 2–3, animals received 10 administrations of 0.3 mg kg⁻¹ GS-986 (Gilead Sciences) by oral gavage every 2 weeks from weeks 50–70. In the combination intervention group, animals began GS-986 treatment at peak immunity 2 weeks after the first MVA boost immunization. TLR7 triggering is known to activate dendritic cells and lymphocytes and to lead to innate immune activation, including secretion of cytokines and chemokines^{16,17}. We observed nonspecific activation of CD8⁺ and CD4⁺ T cells, as measured by CD69 expression 1–2 days following each GS-986 administration (Extended Data Figs 1, 2), and increased plasma levels of IFN- α (Extended Data Fig. 3), thus confirming the immunostimulatory activity of GS-986. Other proinflammatory cytokines and chemokines were also induced by GS-986, including IL-1RA, IL-6, IL-23, CXCL9 (MIG), CXCL11 (I-TAC), CCL4 (MIP-1 β), and CCL11 (Eotaxin) (data not shown).

We next evaluated the immunogenicity of the Ad26/MVA vaccine. The groups that received the vaccine demonstrated a robust >100-fold increase in the magnitude of Gag-, Pol- and Env-specific cellular immune responses, as compared with pre-vaccination responses by IFN- γ ELISPOT assays (Fig. 1b). Cellular immune responses increased substantially at week 28 after priming with Ad26 and further increased at week 50 after administration of MVA boost against both vaccine-matched SIV_{smE543} peptides and virus-matched SIV_{mac239} peptides. We also observed induction of robust Gag-, Pol- and Env-specific CD8⁺ and CD4⁺ T-cell responses, as measured by multiparameter intracellular cytokine staining assays (Extended Data Figs 4, 5). These responses were higher magnitude than those elicited with this same vaccine in SIV-uninfected rhesus monkeys⁴.

Ad26/MVA vaccination also expanded cellular immune breadth by at least 9.2-fold, as measured by IFN- γ ELISPOT assays using subpools of 10 peptides spanning Gag, Pol, and Env (Fig. 1c, Extended Data Fig. 6). Total breadth was defined as the number of Gag-, Pol- and Env-positive subpools. We were unable to fine map individual epitopes owing to insufficient availability of cells given the number of positive subpools. Before vaccination at week 24, we observed an average of 1.9 positive

¹Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA. ²Ragon Institute of MGH, MIT, and Harvard, Cambridge, Massachusetts 02139, USA. ³Program for Evolutionary Dynamics, Harvard University, Cambridge, Massachusetts 02138 USA. ⁴Bioqual, Rockville, Maryland 20852, USA. ⁵Janssen Infectious Diseases and Vaccines, 2301 Leiden, The Netherlands. ⁶Gilead Sciences, Foster City, California 94040, USA. ⁷US Military HIV Research Program, Walter Reed Army Institute of Research, Silver Spring, Maryland 20910, USA. †Present address: International Vaccine Institute, Seoul, South Korea.

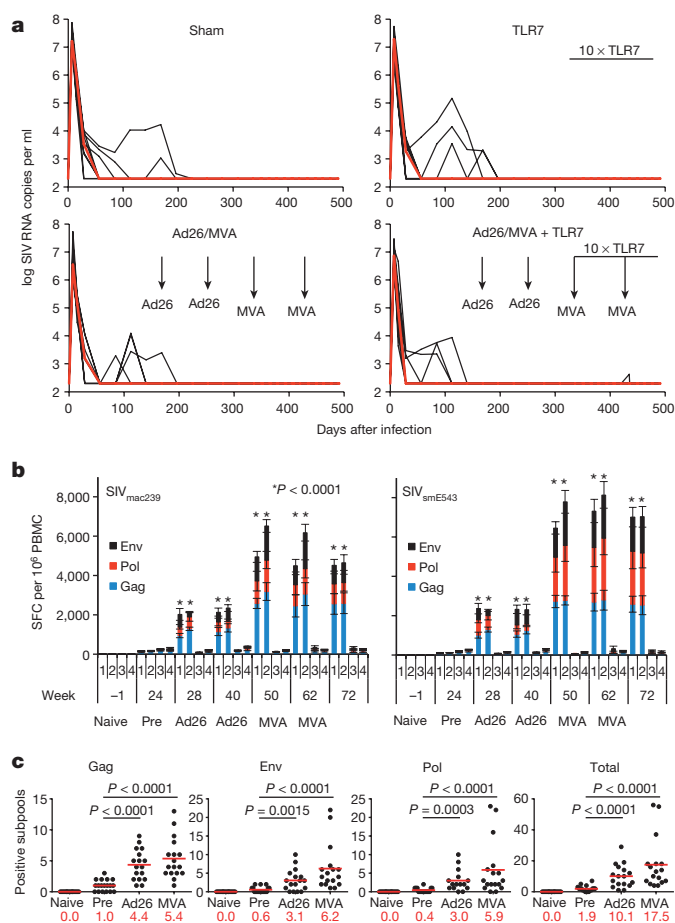


Figure 1 | SIV RNA and vaccine immunogenicity before ART discontinuation. **a**, Rhesus monkeys were infected with SIV_{mac251} on day 0 and began ART on day 7 ($n = 9$ animals per group). Ad26 and MVA vaccination time points are shown with vertical arrows. The timeframe for TLR7 agonist administration (10 doses every 2 weeks) is shown by the horizontal bar. log SIV RNA copies per ml are shown (limit of detection, 2.3 log RNA copies per ml). **b**, IFN- γ ELISPOT responses in response to Gag, Pol and Env peptide pools from SIV_{mac239} and SIV_{smE543}. Group numbers and time points are denoted on the x axis. Group 1 received the Ad26/MVA vaccine alone, and group 2 received the Ad26/MVA vaccine plus TLR7 agonist. P values indicate two-sided Wilcoxon rank-sum tests compared with week 24 (pre-vaccination). **c**, Cellular immune breadth in the vaccinated animals as measured by subpools of 10 peptides spanning the SIV_{mac239} Gag, Pol, and Env proteins, as well as total proteins (Gag+Pol+Env), at week 0 (naive), week 20 (pre-vaccination), week 29 (Ad26), and week 50 (MVA). The numbers of positive subpools are indicated in red. P values indicate two-sided Wilcoxon rank-sum tests.

subpools per animal (1.0 Gag, 0.4 Pol, 0.6 Env). After priming with Ad26, cellular immune breadth markedly expanded to an average of 10.1 positive subpools per animal (4.4 Gag, 3.0 Pol, 3.1 Env). Following MVA boost administration, cellular immune breadth further expanded to an average of 17.5 positive subpools per animal (5.4 Gag, 5.9 Pol, 6.2 Env) (Fig. 1c), which probably represents an underestimate of breadth, as some positive subpools may have contained more than one epitope. Several animals developed Gag–Pol–Env-specific T cells that targeted more than 50 epitopes. The expanded cellular immune breadth involved induction of a large number of new epitopes and did not simply reflect expansion of previously established responses, as the vast majority of the epitopes following vaccination were not observed before vaccination in conventional IFN- γ ELISPOT assays or in assays using peripheral blood mononuclear cells (PBMCs) stimulated *in vitro* with these peptides for enhanced sensitivity (data not shown). We speculate that early initiation of ART induced preserved CD4⁺ T-cell help (Extended Data Fig. 5), which probably contributed to vaccine

immunogenicity. Consistent with this hypothesis is the observation that cellular immune breadth correlated inversely with pre-ART day 7 SIV RNA levels (Extended Data Fig. 7), which may be a surrogate marker for immunologic damage. In contrast with robust cellular immune responses, relatively modest humoral immune responses were observed following vaccination, including binding antibody responses by ELISA and functional antibody-dependent cellular phagocytosis, neutrophil phagocytosis, and natural killer cell activation (data not shown).

We next assessed viral DNA levels in lymph nodes and PBMCs using a reverse transcription (RT)–PCR assay¹⁴ with a sensitivity of 3 DNA copies per 10⁶ CD4⁺ T cells (Fig. 2a, b). In sham controls, viral DNA declined slightly between weeks 20 and 48, presumably reflecting the effect of suppressive ART, but no further decline was observed at week 70. In contrast, the two groups that received the Ad26/MVA vaccine demonstrated marked reductions of viral DNA to undetectable levels in the majority of animals by week 70 in both lymph nodes (Fig. 2a) and PBMCs (Fig. 2b), suggesting that vaccination led to substantial reductions in SIV-infected CD4⁺ T cells in these tissue compartments. It is possible that a larger fraction of proviruses might be intact following ART initiation during acute infection as compared with chronic infection, although this remains to be determined⁸. Viral out-growth assays using 20 million PBMCs were negative in all animals

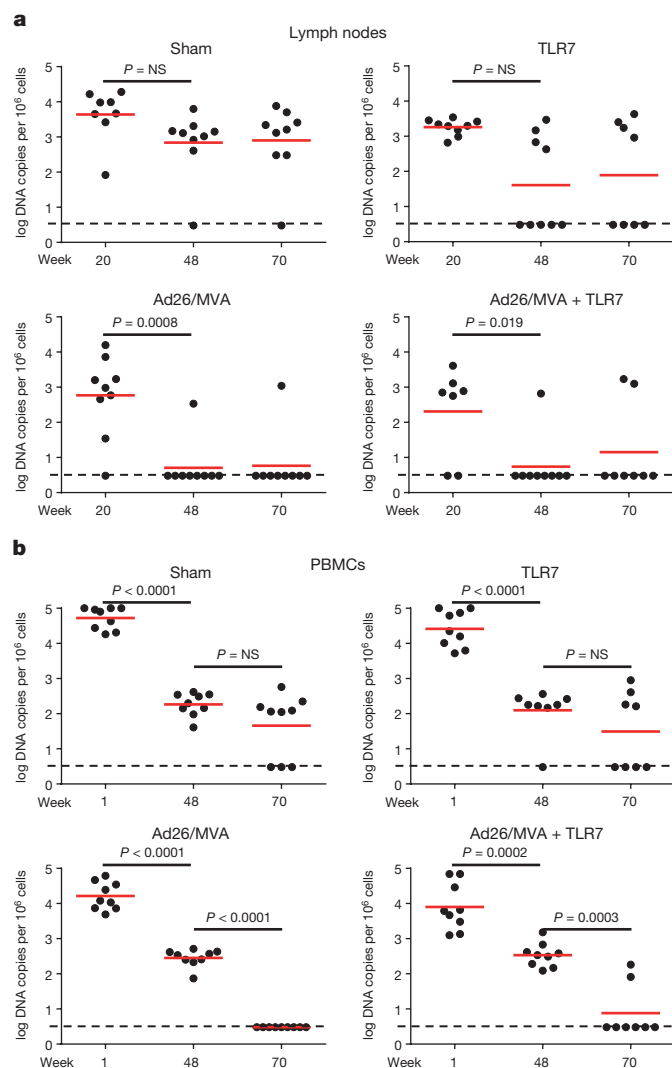


Figure 2 | SIV DNA before ART discontinuation. **a**, **b**, log SIV DNA copies per 10⁶ CD4⁺ T cells are shown (limit of detection, 3 DNA copies per 10⁶ cells) in inguinal lymph node mononuclear cells (**a**) and PBMCs (**b**). Week 48, after Ad26 prime; week 70 after MVA boost. P values indicate two-sided Wilcoxon rank-sum tests.

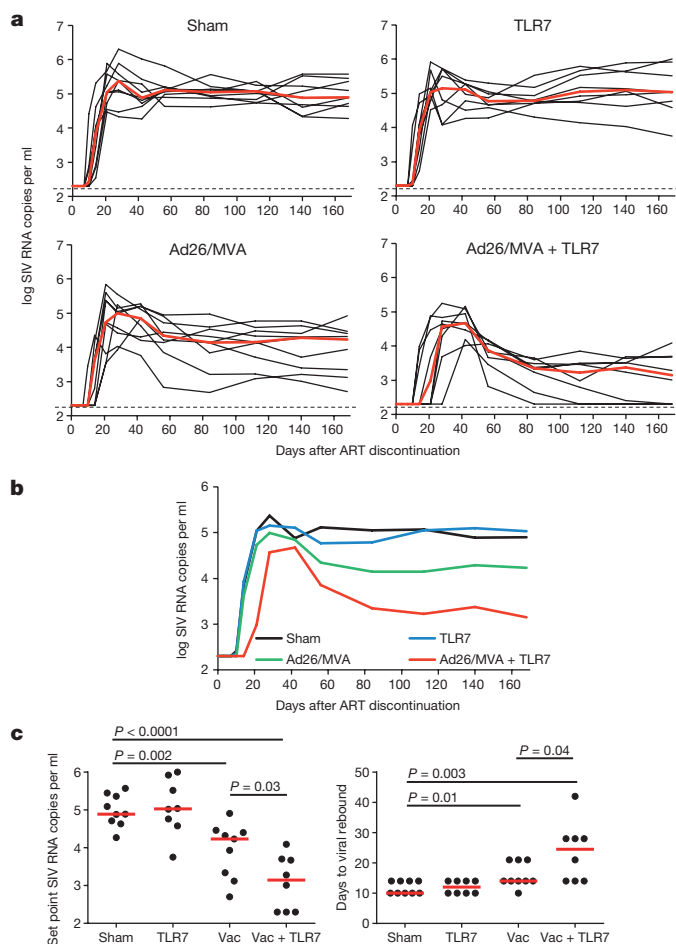


Figure 3 | SIV RNA following ART discontinuation. **a**, log SIV RNA copies per ml are shown (limit of detection, 2.3 log RNA copies per ml) following ART discontinuation at study week 72. Days following ART discontinuation are shown on the x axis. **b**, Median log SIV RNA in each group. **c**, Statistical analysis of set point levels of SIV RNA and time to viral rebound in each group. Vac, vaccination with Ad26/MVA. P values indicate two-sided Wilcoxon rank-sum tests.

including controls at week 70 (data not shown), presumably as a result of early initiation of ART.

To evaluate the therapeutic efficacy of the interventions, we discontinued ART at week 72. Viral rebound was observed in all animals (Fig. 3a). All sham controls rebounded by day 10–14 following ART discontinuation in a stereotypical fashion and exhibited median set point plasma SIV RNA levels of 4.89 log copies per ml (range 4.27–5.57 log copies per ml) on day 168 following ART discontinuation. The monkeys that received GS-986 alone did not demonstrate any discernible delay or control of viral rebound, indicating that TLR7 stimulation alone exerted no detectable antiviral effect in this study. Animals that received the Ad26/MVA vaccine alone exhibited a 0.66 log reduction of median set point plasma SIV RNA levels to 4.23 log copies per ml (range 2.70–4.91 log copies per ml) ($P = 0.002$, Wilcoxon rank-sum test) but only a marginal delay of viral rebound ($P = 0.01$, Wilcoxon rank-sum test), as compared with controls (Fig. 3b, c). By contrast, monkeys that received both the Ad26/MVA vaccine and GS-986 showed a notable 1.74 log reduction of median set point plasma SIV RNA levels to 3.15 log copies per ml (range <2.30–4.09 log copies per ml) ($P < 0.0001$) and a 2.5-fold delay of viral rebound from a median of 10 to 25 days, as compared with controls ($P = 0.003$) (Fig. 3b, c). Moreover, 33% (3 of 9) of the monkeys in the combination intervention group showed effective virologic control to undetectable set point viral loads (<2.30 log copies per ml) following ART discontinuation. These data demonstrate that the combination of Ad26/MVA vaccination and

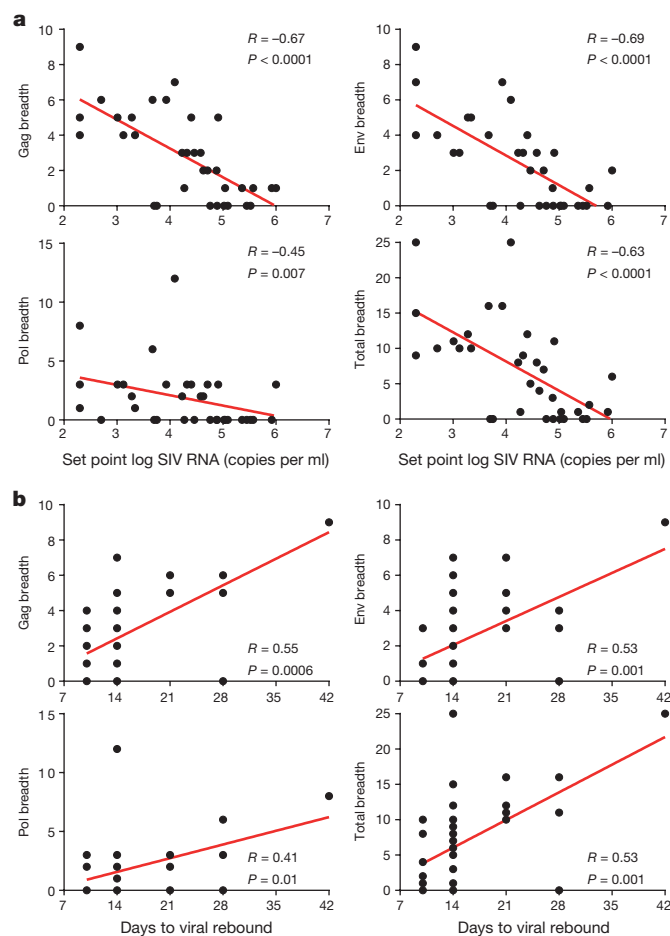


Figure 4 | Correlations of cellular immune breadth with set point viral loads and time to viral rebound. **a**, **b**, Correlations are shown for the breadth of Gag, Pol, Env and total (Gag+Pol+Env) cellular immune responses, as defined as the number of positive subpools at the time of ART discontinuation at week 72 and set point log SIV RNA (**a**) or time to viral rebound (**b**) following ART discontinuation. P values indicate two-sided Spearman rank-correlation tests. R values indicate correlation coefficients.

TLR7 stimulation improved virologic control and delayed viral rebound following ART discontinuation.

We next evaluated the immunologic and virologic correlates of virologic control. Cellular immune breadth immediately before ART discontinuation (Fig. 4a) and at peak immunity (Extended Data Fig. 8) correlated inversely with set point viral loads following ART discontinuation, particularly the breadth of Gag, Env, and total cellular immune responses ($P < 0.0001$, Spearman rank-correlation tests). Gag, Env, and total cellular immune breadth also correlated directly with the time to viral rebound (Fig. 4b, Extended Data Fig. 9; $P = 0.0001$ to $P = 0.001$). Consistent with the correlates analyses, mathematical modelling of viral dynamics further suggested that the combination of changes in the reservoir exit rate and the early viral growth rate accounted for the differences in the time to viral rebound, whereas the virus-specific immune proliferation rate was probably responsible for virologic control (Extended Data Fig. 10).

Viral DNA levels in lymph nodes and PBMCs correlated poorly with virologic control following ART discontinuation and time to viral rebound ($P = 0.03$, data not shown), presumably because all animals with undetectable viral DNA still experienced rebound. These findings suggest that these viral DNA assays are not sufficiently sensitive to predict functional cure, consistent with recent clinical observations^{18,19}. Moreover, there was no correlation between day 7 pre-ART SIV RNA levels or time to initial virologic suppression on ART and virologic control following ART discontinuation (P value not significant).

In this study, we demonstrate that Ad26/MVA therapeutic vaccination robustly augmented cellular immune magnitude and breadth in ART-suppressed, SIV-infected rhesus monkeys and that the TLR7 agonist GS-986 led to innate immune stimulation and cellular activation. The combination of Ad26/MVA vaccination and GS-986 resulted in a significant 1.74 log reduction in median set point viral loads and a 2.5-fold delay in the time to viral rebound following ART discontinuation, as compared with sham controls. Moreover, 3 of 9 animals in this group demonstrated virologic control to undetectable levels in the absence of ART. These three animals were characterized by high cellular immune magnitude and breadth and negative viral DNA before ART discontinuation. Taken together, these data demonstrate the proof-of-concept that the combination of therapeutic vaccination and innate immune stimulation can affect viral rebound following ART discontinuation.

The Ad26/MVA vaccine induced remarkably potent IFN- γ ELISPOT responses in the ART-suppressed, SIV-infected monkeys in the present study, consistent with a previous prophylactic vaccine study in uninfected monkeys⁴. We speculate that preserved CD4⁺ T-cell help, primed by short-term viral replication following SIV infection before initiation of ART on day 7, probably enhanced vaccine immunogenicity. The Ad26/MVA vaccine also expanded cellular immune breadth by nearly tenfold, including induction of responses to a large number of epitopes that were not detectable after SIV infection. This may be critical for a therapeutic vaccine, as the viral reservoir typically contains viruses with T-cell-epitope escape mutations¹². We did not detect evidence for viral 'blipping' on day 1 or day 2 following TLR7 agonist administration (J. Whitney, unpublished data), which may reflect the early initiation of ART on day 7 of infection and thus the limited size of the viral reservoirs in this study¹⁴. We were therefore unable to determine whether the beneficial effect of the TLR7 agonist reflected its potential role as a vaccine adjuvant, a latency reversing agent, or both. Future studies should be performed using SIV-infected monkeys for whom ART is initiated during chronic infection, as these would be more representative of the majority of HIV-1-infected individuals. Moreover, future studies could explore longer periods of ART suppression, potentially to reduce residual viral replication, although the majority of animals appear to have a stable reservoir after 24–72 weeks of ART¹⁴. The capacity of immunologic interventions to target follicular helper CD4⁺ T cells in lymph nodes should also be explored²⁰.

Previous studies of poxvirus-, adenovirus- and DNA-based therapeutic vaccines have typically shown only a modest effect on viral rebound following ART discontinuation in both rhesus monkeys²¹ and humans^{22,23}. The present study extends these previous observations by combining therapeutic vaccination with innate immune stimulation. Of note, the combination of Ad26/MVA vaccination and TLR7 stimulation proved more potent than either component alone. This finding is consistent with a previous *in vitro* study that showed that robust CD8⁺ T cells may be able to facilitate elimination of the viral reservoir following reactivation¹¹. The present study also demonstrates that the breadth of Gag-, Pol- and Env-specific T-cell responses correlated inversely with set point viral loads following ART discontinuation, suggesting that the mechanism underlying the therapeutic efficacy of the vaccine involved expansion of cellular immune breadth and immunologic control of virus rebounding from the reservoir.

In summary, our data suggest the potential of combining therapeutic vaccination with innate immune stimulation as an HIV-1 cure strategy. Our findings show that these interventions can improve virologic control and delay viral rebound following ART discontinuation in SIV-infected rhesus monkeys that began ART during acute infection. Additional preclinical and clinical studies with Ad26/MVA vaccination and TLR7 stimulation should be performed to explore this strategy in greater detail.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 September; accepted 1 November 2016.

Published online 9 November 2016.

- Barouch, D. H. & Deeks, S. G. Immunologic strategies for HIV-1 remission and eradication. *Science* **345**, 169–174 (2014).
- Deeks, S. G. *et al.* International AIDS Society global scientific strategy: towards an HIV cure 2016. *Nat. Med.* **22**, 839–850 (2016).
- Barouch, D. H. *et al.* Protective efficacy of a global HIV-1 mosaic vaccine against heterologous SHIV challenges in rhesus monkeys. *Cell* **155**, 531–539 (2013).
- Barouch, D. H. *et al.* Vaccine protection against acquisition of neutralization-resistant SIV challenges in rhesus monkeys. *Nature* **482**, 89–93 (2012).
- Finzi, D. *et al.* Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* **278**, 1295–1300 (1997).
- Persaud, D., Zhou, Y., Siliciano, J. M. & Siliciano, R. F. Latency in human immunodeficiency virus type 1 infection: no easy answers. *J. Virol.* **77**, 1659–1665 (2003).
- Chun, T. W. *et al.* Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl Acad. Sci. USA* **94**, 13193–13197 (1997).
- Ho, Y. C. *et al.* Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540–551 (2013).
- Finzi, D. *et al.* Latent infection of CD4⁺ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat. Med.* **5**, 512–517 (1999).
- Chun, T. W., Davey, R. T., Jr, Engel, D., Lane, H. C. & Fauci, A. S. Re-emergence of HIV after stopping therapy. *Nature* **401**, 874–875 (1999).
- Shan, L. *et al.* Stimulation of HIV-1-specific cytolytic T lymphocytes facilitates elimination of latent viral reservoir after virus reactivation. *Immunity* **36**, 491–501 (2012).
- Deng, K. *et al.* Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. *Nature* **517**, 381–385 (2015).
- Liu, J. *et al.* Immune control of an SIV challenge by a T-cell-based vaccine in rhesus monkeys. *Nature* **457**, 87–91 (2009).
- Whitney, J. B. *et al.* Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature* **512**, 74–77 (2014).
- Abbink, P. *et al.* Comparative seroprevalence and immunogenicity of six rare serotype recombinant adenovirus vaccine vectors from subgroups B and D. *J. Virol.* **81**, 4654–4663 (2007).
- Kawai, T. *et al.* Interferon- α induction through Toll-like receptors involves a direct interaction of IRF7 with MyD88 and TRAF6. *Nat. Immunol.* **5**, 1061–1068 (2004).
- Hemmi, H. *et al.* Small anti-viral compounds activate immune cells via the TLR7 MyD88-dependent signaling pathway. *Nat. Immunol.* **3**, 196–200 (2002).
- Henrich, T. J. *et al.* Antiretroviral-free HIV-1 remission and viral rebound after allogeneic stem cell transplantation: report of 2 cases. *Ann. Intern. Med.* **161**, 319–327 (2014).
- Persaud, D. *et al.* Absence of detectable HIV-1 viremia after treatment cessation in an infant. *N. Engl. J. Med.* **369**, 1828–1835 (2013).
- Banga, R. *et al.* PD1⁺ and follicular helper T cells are responsible for persistent HIV-1 transcription in treated aviremic individuals. *Nat. Med.* **22**, 754–761 (2016).
- Hel, Z. *et al.* Viremia control following antiretroviral treatment and therapeutic immunization during primary SIV251 infection of macaques. *Nat. Med.* **6**, 1140–1146 (2000).
- Rosenberg, E. S. *et al.* Safety and immunogenicity of therapeutic DNA vaccination in individuals treated with antiretroviral therapy during acute/early HIV-1 infection. *PLoS One* **5**, e10555 (2010).
- Li, J. Z. *et al.* Factors associated with viral rebound in HIV-1-infected individuals enrolled in a therapeutic HIV-1 gag vaccine trial. *J. Infect. Dis.* **203**, 976–983 (2011).

Acknowledgements We thank C. Linde, T. Broge, T. Barnes, D. van Manen, F. Wegmann, C. Shaver, W. Wagner, M. Boyd, R. Nityanandam, K. Smith, S. Blackmore, L. Parenteau, P. Giglio, M. Shetty, S. Levin, J. Shields, G. Neubauer, and F. Stephens for generous advice, assistance, and reagents. We acknowledge support from the US Army Medical Research and Materiel Command and the Military HIV Research Program, Walter Reed Army Institute of Research through its cooperative agreement with the Henry M. Jackson Foundation (W81XWH-11-2-0174); the National Institutes of Health (AI096040, AI124377, AI126603, OD019851); the Ragon Institute of MGH, MIT, and Harvard. Mathematical model fitting was performed on the Orchestra High Performance Compute Cluster at Harvard Medical School. The views expressed in this manuscript are those of the authors and do not represent the official views of the Department of the Army or the Department of Defense.

Author Contributions D.H.B., N.L.M., J.H.K., M.L.R., M.G.P., H.S., and R.G. designed the studies. J.H. and R.G. developed the ART formulation and TLR7 agonist. E.N.B., C.C., K.E.S., J.L., J.P.N., A.L.B., L.P., B.C.L., J.J., D.J., J.M., S.M., A.C., K.M., and G.A. performed the immunologic assays. P.A. and D.N. conducted the virologic assays. J.M.G. and A.L.H. performed the viral dynamics modelling. M.G.L. led the clinical care of the rhesus monkeys. D.H.B. wrote the paper with all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.H.B. (dbarouch@bidmc.harvard.edu).

Reviewer Information Nature thanks S. Lewin and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Animals. 36 outbred Indian-origin, young adult male and female rhesus monkeys (*Macaca mulatta*) were genotyped and selected as negative for the protective MHC class I alleles *Mamu-A*01*, *Mamu-B*08*, and *Mamu-B*17*. Animals expressing susceptible and resistant TRIM5 α alleles were distributed among the groups. Animals were otherwise randomly allocated to groups. All monkeys were housed at Bioqual, Rockville, Maryland. Animals were infected with 500 TCID₅₀ (50% tissue culture infective dose) of our SIV_{mac251} challenge stock^{4,13,24} by the intrarectal route. Monkeys were bled up to two times per week for viral load determinations. Immunologic and virologic assays were performed blinded. All animal studies were approved by the appropriate Institutional Animal Care and Use Committee (IACUC). The sample size was estimated to achieve 80% power to detect a 1.0 log difference in setpoint viral loads.

ART regimen. The preformulated antiretroviral therapy (ART) cocktail contained 5.1 mg ml⁻¹ tenofovir disoproxil fumarate (TDF), 40 mg ml⁻¹ emtricitabine (FTC), and 2.5 mg ml⁻¹ dolutegravir (DTG) in a solvent containing 15% (v/v) kleptose adjusted to pH 4.2. This ART cocktail was administered once daily at 1 ml kg⁻¹ body weight via the subcutaneous route.

Ad26/MVA vaccination and TLR7 agonist GS-986 administration. In groups 1–2, monkeys were primed by the intramuscular route with 3 × 10¹⁰ viral particles of Ad26 vectors^{4,15} expressing SIV_{smE543} gag-pol-env at weeks 24 and 36, and were boosted with 10⁸ plaque-forming units of MVA vectors⁴ expressing SIV_{smE543} gag-pol-env at weeks 48 and 60. In groups 2–3, animals received 10 administrations of 0.3 mg per kg GS-986 (Gilead Sciences) by oral gavage every 2 weeks from weeks 50–70.

Cellular immune assays. SIV-specific cellular immune responses were assessed by IFN- γ ELISPOT assays and multiparameter intracellular cytokine staining assays essentially as described⁴. Estimates of cellular immune breadth involved IFN- γ ELISPOT assays using subpools of 10 peptides across the Gag, Pol, and Env proteins. 12-colour intracellular cytokine staining assays were performed with the Aqua green-fluorescent reactive dye (Invitrogen, L23101) and predetermined titres of monoclonal antibodies (Becton Dickinson) against CD3 (SP34; Alexa Fluor 700), CD4 (OKT4; BV711, Biolegend), CD8 (SK1; allophycocyanin-cyanine 7 (APC-Cy7)), CD28 (L293; BV610), CD95 (DX2; allophycocyanin (APC)), CD69 (TP1.55.3; phycoerythrin-Texas red (energy-coupled dye; ECD); Beckman Coulter), gamma interferon (IFN- γ) (B27; phycoerythrin-cyanine 7 (PE-Cy7)), Ki67 (B56; fluorescein isothiocyanate (FITC)), CCR5 (3A9; phycoerythrin (PE)), CCR7 (3D12; Pacific Blue), and PD-1 (EH21.1; peridinin chlorophyll-A-cyanine 5.5 (PerCP-Cy5.5)). IFN- γ backgrounds were consistently <0.01% in PBMCs.

Viral RNA assays. Viral RNA was isolated from cell-free plasma using a viral RNA extraction kit (Qiagen) and was quantitated essentially as described¹⁴.

Viral DNA assays. Levels of proviral DNA were quantitated as previously described¹⁴. Total cellular DNA was isolated from 5 × 10⁶ cells using a QIAamp DNA Blood Mini kit (Qiagen). The absolute quantification of viral DNA in each sample was determined by quantitative PCR (qPCR) using primers specific to a conserved region of SIV_{mac239}. All samples were directly compared to a linear virus standard and the simultaneous amplification of a fragment of human GAPDH gene. PCR assays were performed with 100–200 ng sample DNA.

Statistical analyses. Analysis of virologic and immunologic data was performed using GraphPad Prism v6.03 (GraphPad Software). Comparisons of groups was performed using two-sided Wilcoxon rank-sum tests without Bonferroni adjustments. Correlations were assessed by two-sided Spearman rank-correlation tests without Bonferroni adjustments.

Viral dynamics modelling. We employed viral dynamics modelling to better characterize the kinetics of viral rebound and to gain insight into the mechanism behind the improved control of rebound in certain groups of treated animals. We were particularly interested in whether the effect of the vaccine and TLR7 agonist on rebound could be explained by a reduction in the latent reservoir size, or an antiviral immune response, or both.

Most existing viral dynamics models for HIV-1/SIV or other infections are not able to explain the diversity of kinetics seen in different animals in this study—such as rebound to a high set point, rebound to a high peak and lower set point, and rebound to a medium or high peak followed by control to undetectable levels—with mechanisms that are realistic for HIV-1/SIV. Specifically, for HIV-1 and SIV infection, exhaustion of target cells cannot alone explain the decline from peak viraemia, which instead requires an immune response that can markedly reduce the viral load set point without significant changes in the death rate of virus-producing infected cells²⁵. On the basis of other modelling work^{26–29}, we developed a new model to explain all of these observations.

The model we used is described by a system of ordinary differential equations that track changes in the levels of uninfected (*T*) and infected (*I*) target cells, free virus (*V*), and precursor (*P*) and effector (*E*) immune responses over time (Extended Data Fig. 10):

$$\dot{T} = \lambda - \beta TV - d_T T$$

$$\dot{I} = a + \frac{\beta TV}{1 + (E/N_E)} - d_I I$$

$$\dot{V} = kI - cV$$

$$\dot{P} = m + p(1 - f) \frac{V}{V + N_P} P - d_P P$$

$$\dot{E} = pf \frac{V}{V + N_P} P - d_E E$$

All variables are expressed as concentrations per ml of plasma. Parameter λ is the rate of production of susceptible uninfected cells (cells per ml per day), β is the viral infectivity (per (copies per ml) per day), d_I is the death rate of infected cells (per day), d_T is the death rate of uninfected cells (per day), a is the rate at which latent cells reactivate to become productively infected (cells per day), k is the viral burst size (virions per cell per day) and c is the viral clearance rate (per day). Our model of the immune response is general enough to describe either humoral or cellular effects. Precursor immune cells (naive and memory) are produced at a baseline rate m (cells per ml per day), proliferate at an antigen-dependent rate, and die at a rate d_P (per day). The maximum proliferate rate is p (per day) and half-maximal proliferation occurs at viral load N_P (copies per ml). A fraction, f , of all proliferating cells differentiate into effectors, which die at a rate d_E (per day). Effectors reduce the rate at which actively infected cells are produced (either by inactivating free virus or killing early-stage infected cells^{29,30}), with half-maximal inhibition occurring at a concentration N_E (cell per ml).

During ART, $\beta = 0$, and virus and cells reach steady states at values $T_0 = \lambda/d_T$, $I_0 = a/d_I$, $V_0 = (a/d_I)(k/c)$, $P_0 \approx (m/d_P)$, and $E_0 \approx 0$ (as $V_0 \ll N_P$), which we take as the initial conditions at the time of ART interruption.

As we only have longitudinal observations of viral load, we cannot uniquely identify all the parameters of this model. We conducted extensive analytic and numeric investigation of the model to determine which parameters were most important for the dynamics and which could be estimated from the available data. On the basis of this knowledge and our previous work¹⁴, we fixed the parameters $d_I = 0.4$ (per day)¹⁴, $d_T = 0.05$ (per day), $d_P = 0.001$ (per day), $d_E = 1$ (per day), $f = 0.9$. As only the ratio m/N_P can be identified from the viral load data, we fixed $N_E = 10^4$. The ratio k/c was fixed to 2,000 viruses per cell based on a burst size of $k = 5 \times 10^4$ (virions per cell per day) and $c = 23$ per day for SIV^{31–33}. The other parameters (λ , β , a , m , p , N_P) were fitted. We also fitted results to a simpler model with no antigen-dependent immune response (P , $E = 0$) to ensure that the fits were improved by adding in the extra terms. We modelled the effect of antiretroviral drug decay by including a three-day washout period, during which the virus could not productively reproduce. Varying this period between 0 and 6 days did not change conclusions from the model.

Models were fit to each animal individually using a Bayesian framework. Briefly, a joint posterior distribution was estimated for the fitted parameters using a likelihood function for the viral load at each time point and very weakly informative priors on the fitted parameters (see below). We parameterized the model using log-transformed parameters to account for the possibility of large parameter variations between monkeys. The likelihood assumes that the observed viral load is log-normally distributed around the true viral load, and that measurements include an error with variance σ^2 (which we fixed at 0.1 based on previous work). The estimated posterior distribution was sampled using the Metropolis–Hastings algorithm implemented in R using the package MHadaptive³⁴ and the differential equations were numerically integrated using deSolve³⁵. After a burn-in of 10⁴ iterations, 10⁵ updates were run and 1 in 10 samples were collected. The resulting fits were visually inspected to make sure they were consistent with the observed viral load. We specified the prior for each parameter independently according to:

$$\log_{10}(p_i) \sim \text{Laplace}(\mu_i, 1)$$

with the means for each parameter $\log_{10}(\lambda) = 1.3$, $\log_{10}(\beta) = -6$, $\log_{10}(a) = -4.5$, $\log_{10}(m) = 0$, $\log_{10}(p) = -4$, and $\log_{10}(N_P) = 4$.

Beyond the individual parameters, we also calculated a composite parameter describing the exponential growth rate of the virus immediately following rebound, r , as

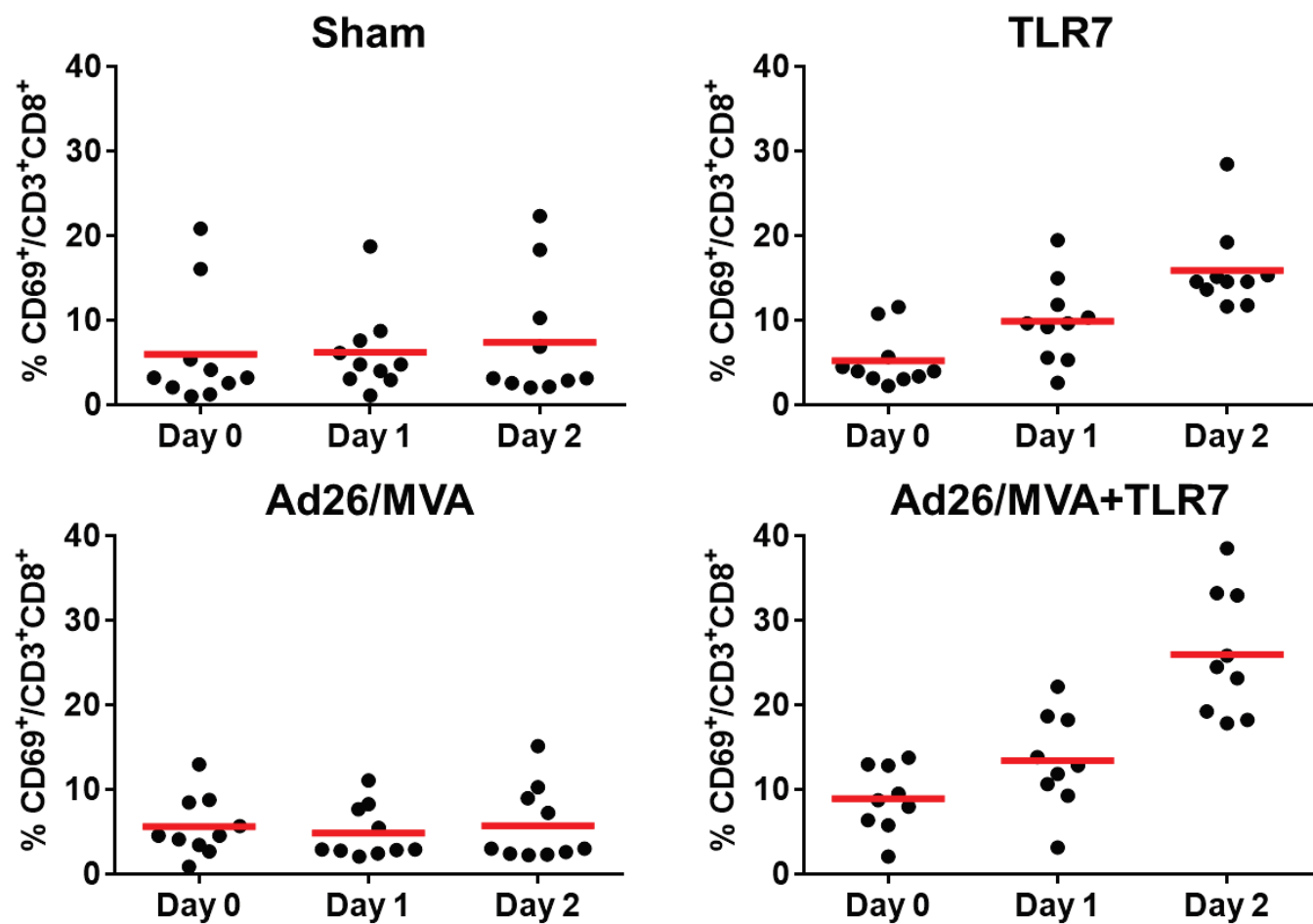
$$r = \frac{\lambda\beta}{d_T} \left(\frac{k}{c} \right) - d_I$$

We also calculated the time of rebound (τ_r) as the time at which viral load was expected to be exactly equal to the detection limit of $v_d = 200$ copies per ml by numerically integrating the differential equations until $V(t) = v_d$.

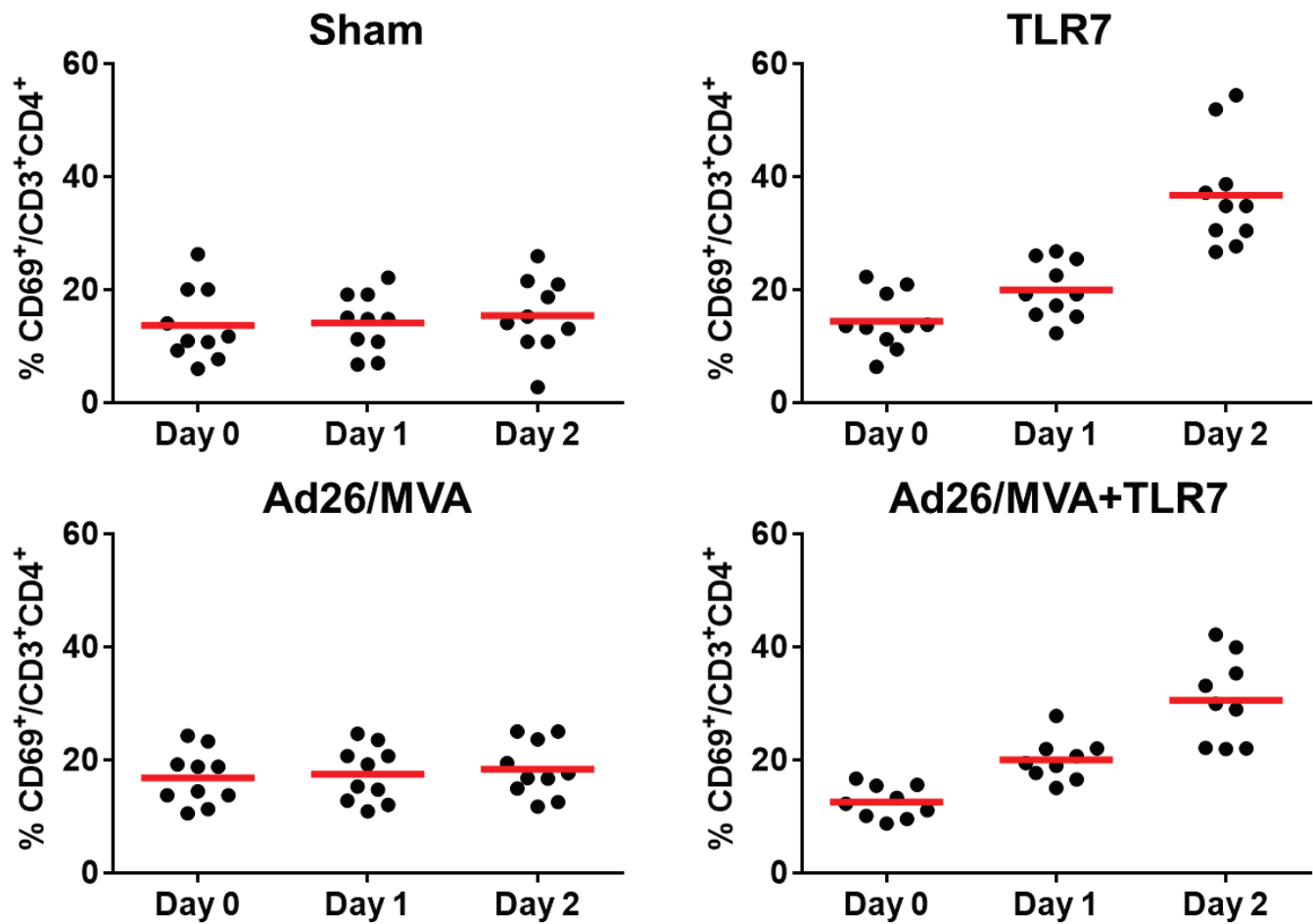
In Extended Data Fig. 10, we show results of the fits. Values shown represent the median and 95% credible intervals of the posterior of the mean of the group. The posterior mean was constructed by repeatedly sampling one value from the posterior of each animal in the group, and then taking the mean. *P* values reported for an increase in the posterior mean represent the probability that a random sample from the posterior mean in one group is smaller than a sample from the posterior mean in the other.

Data availability. All data generated and analysed in this study are available from the corresponding author upon reasonable request.

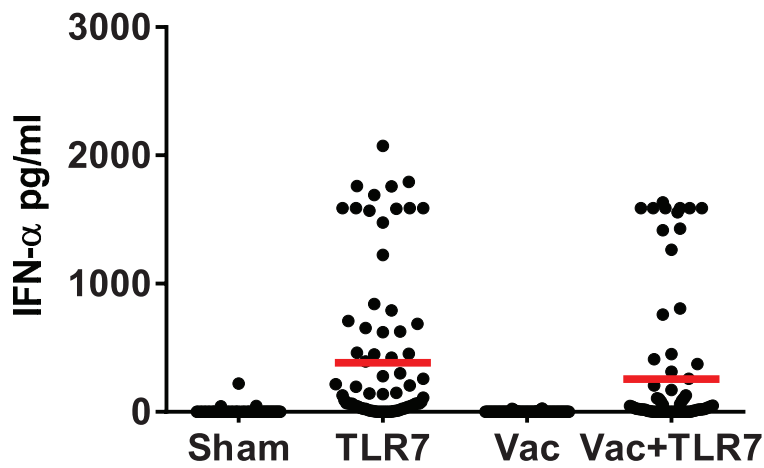
24. Liu, J. *et al.* Low-dose mucosal simian immunodeficiency virus infection restricts early replication kinetics and transmitted virus variants in rhesus monkeys. *J. Virol.* **84**, 10406–10412 (2010).
25. Burg, D., Rong, L., Neumann, A. U. & Dahari, H. Mathematical modeling of viral kinetics under immune control during primary HIV-1 infection. *J. Theor. Biol.* **259**, 751–759 (2009).
26. Wodarz, D. *et al.* A new theory of cytotoxic T-lymphocyte memory: implications for HIV treatment. *Philos. Trans. R. Soc. B Biol. Sci.* **355**, 329–343 (2000).
27. Luzyanina, T., Engelborghs, K., Ehl, S., Klenerman, P. & Bocharov, G. Low level viral persistence after infection with LCMV: a quantitative insight through numerical bifurcation analysis. *Math. Biosci.* **173**, 1–23 (2001).
28. Althaus, C. L. & De Boer, R. J. Dynamics of immune escape during HIV/SIV infection. *PLOS Comput. Biol.* **4**, e1000103 (2008).
29. Elemans, M. *et al.* Why don't CD8⁺ T cells reduce the lifespan of SIV-infected cells *in vivo*? *PLOS Comput. Biol.* **7**, e1002200 (2011).
30. Gadhamsetty, S., Coorens, T. & Boer, R. J. de. Notwithstanding circumstantial alibis, cytotoxic T cells can be major killers of HIV-1 infected cells. *J. Virol.* **306**–316 (2016).
31. Chen, H. Y., Di Mascio, M., Perelson, A. S., Ho, D. D. & Zhang, L. Determination of virus burst size *in vivo* using a single-cycle SIV in rhesus macaques. *Proc. Natl Acad. Sci. USA* **104**, 19079–19084 (2007).
32. De Boer, R. J., Ribeiro, R. M. & Perelson, A. S. Current estimates for HIV-1 production imply rapid viral clearance in lymphoid tissues. *PLOS Comput. Biol.* **6**, e1000906 (2010).
33. Ramratnam, B. *et al.* Rapid production and clearance of HIV-1 and hepatitis C virus assessed by large volume plasma apheresis. *Lancet* **354**, 1782–1785 (1999).
34. Chivers, C. *MHadaptive: General Markov Chain Monte Carlo for Bayesian Inference using adaptive Metropolis-Hastings sampling*. <https://cran.r-project.org/web/packages/MHadaptive/MHadaptive.pdf> (2012).
35. Soetaert, K., Petzoldt, T. & Setzer, R. W. *deSolve: Solvers for Initial Value Problems of Differential Equations (ODE, DAE, DDE)*. <https://cran.r-project.org/web/packages/deSolve/vignettes/deSolve.pdf> (2016).



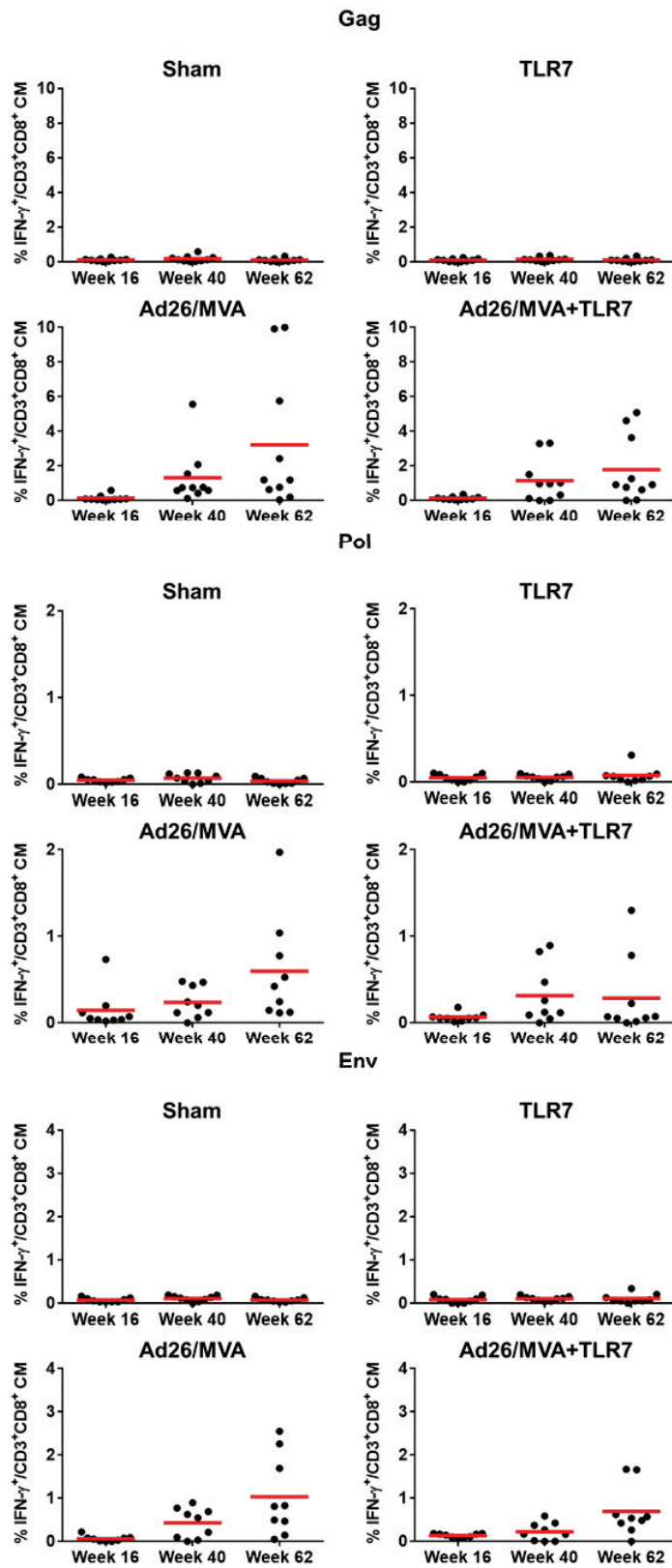
Extended Data Figure 1 | Activation of CD8⁺ T cells following GS-986 administration. Representative data on days 0, 1, and 2 following GS-986 administration. Activation was assessed by CD69 expression on CD3⁺CD8⁺ T cells. Red bars represent mean values for each group.



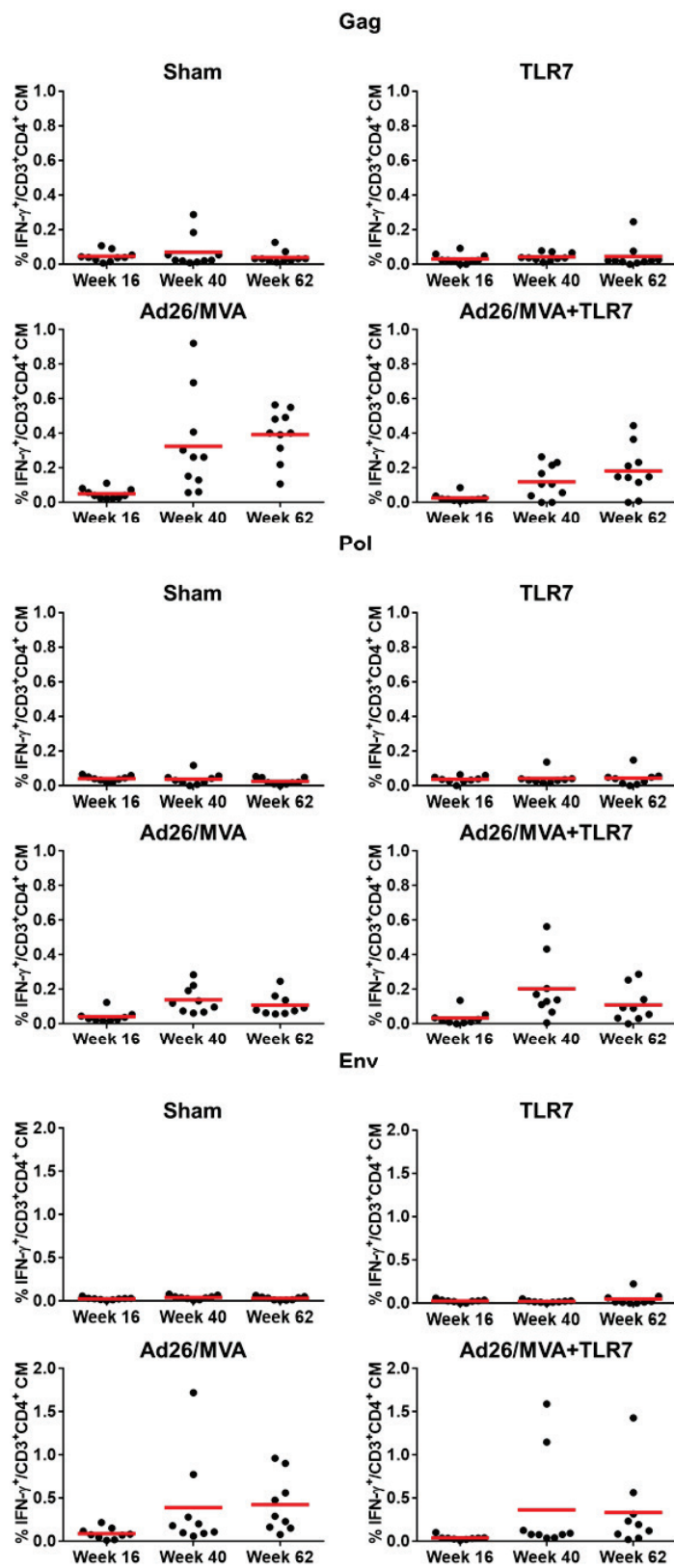
Extended Data Figure 2 | Activation of CD4⁺ T cells following GS-986 administration. Representative data on days 0, 1, and 2 following GS-986 administration. Activation was assessed by CD69 expression on CD3⁺CD4⁺ T cells.



Extended Data Figure 3 | Innate immune stimulation following GS-986 administration. Plasma IFN- α (pg ml^{-1}) levels are shown on day 1 following GS-986 administration. Red bars represent mean values for each group. Data points reflect all animals following all GS-986 administrations combined with pre-dose levels subtracted.

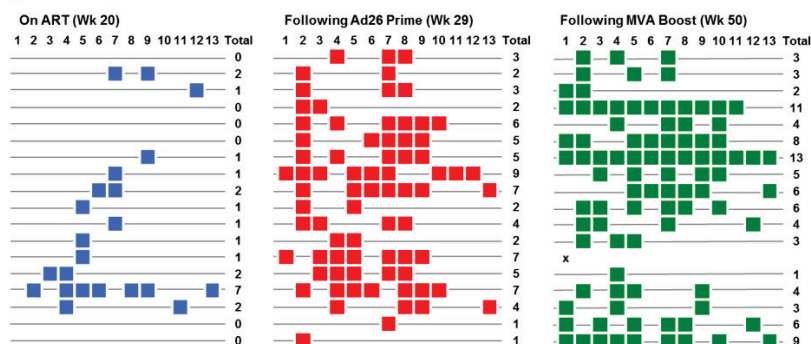


Extended Data Figure 4 | CD8⁺ T cells following Ad26/MVA vaccination.
SIV_{mac239} Gag-, Pol- and Env-specific, IFN- γ ⁺CD3⁺CD8⁺ central memory T cells were assessed by multiparameter intracellular cytokine staining assays.

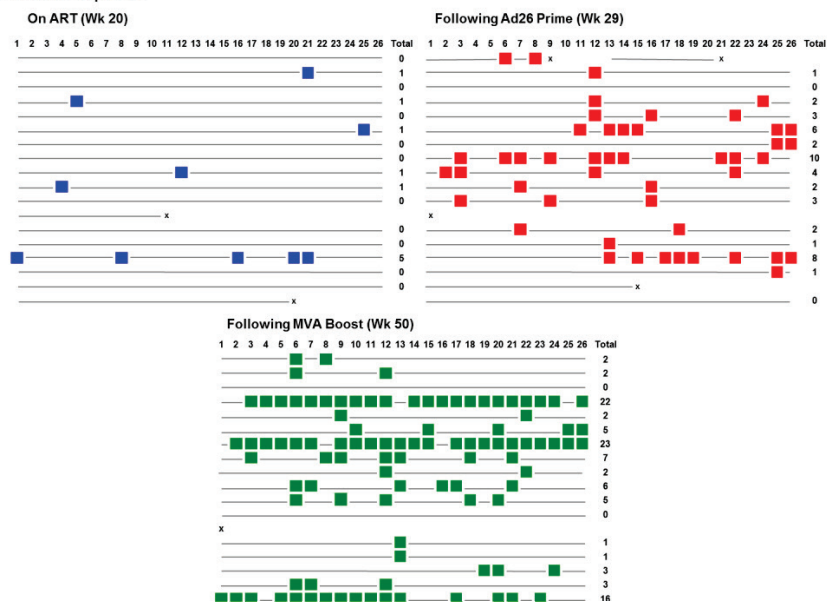


Extended Data Figure 5 | CD4⁺ T cells following Ad26/MVA vaccination. SIV_{mac239} Gag-, Pol- and Env-specific, IFN- γ ⁺CD3⁺CD4⁺ central memory T cells were assessed by multiparameter intracellular cytokine staining assays.

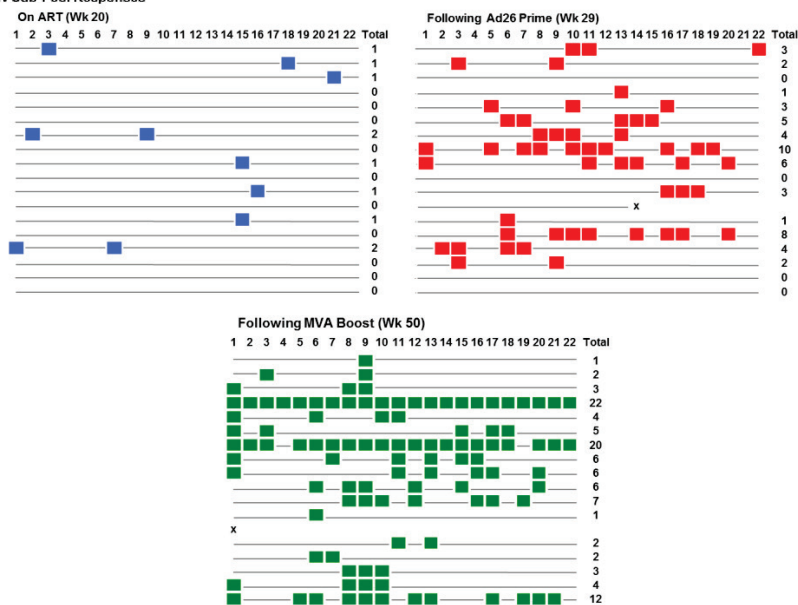
Gag Sub-Pool Responses



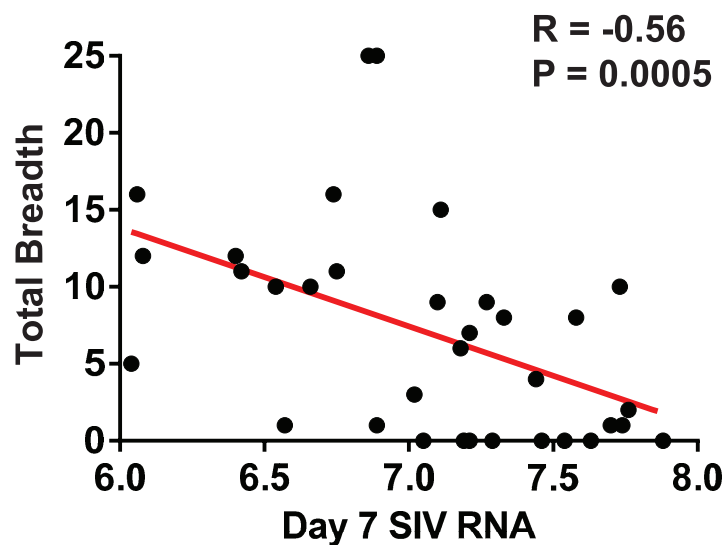
Pol Sub-Pool Responses



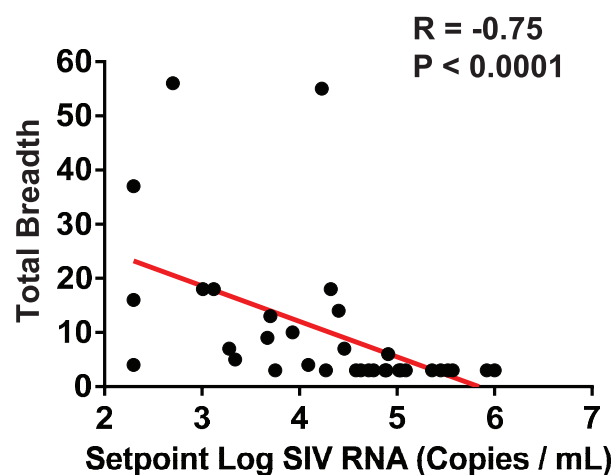
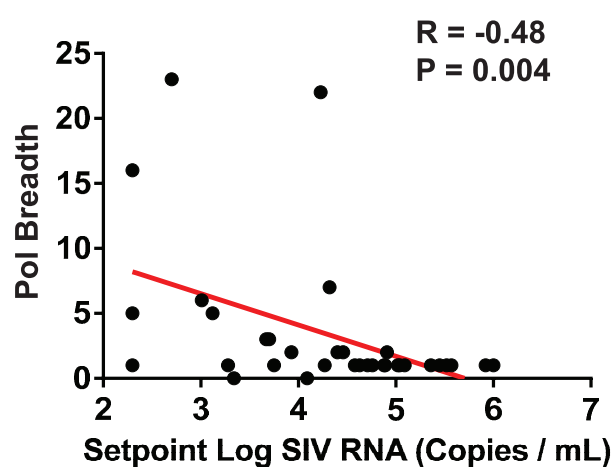
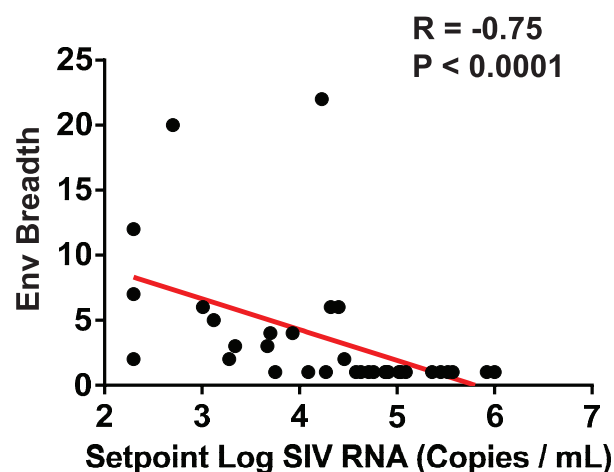
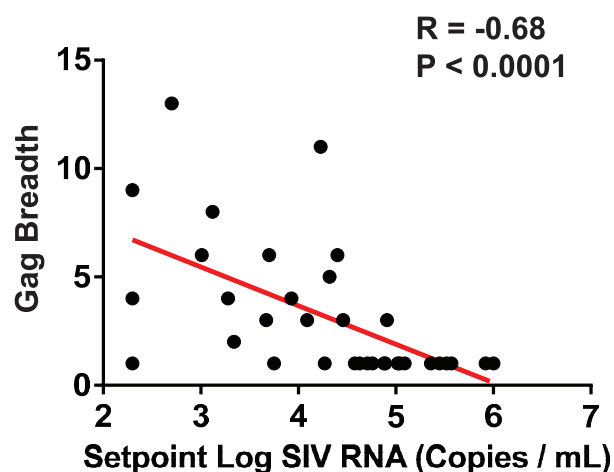
Env Sub-Pool Responses



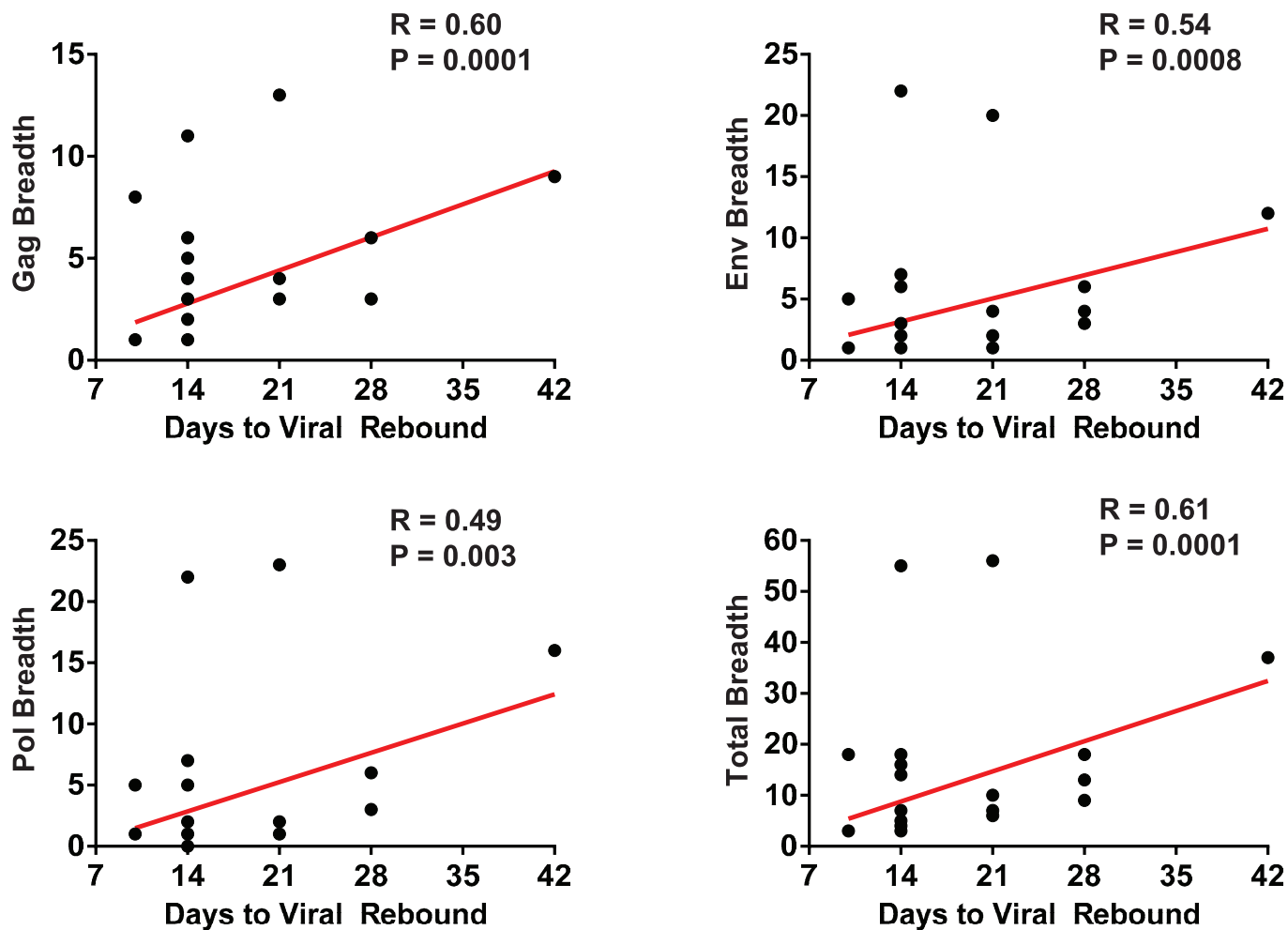
Extended Data Figure 6 | Cellular immune breadth. Responses to subpools of 10 peptides spanning SIV_{mac239} Gag, Pol, and Env are shown before vaccination (week 20, blue), after Ad26 prime (week 29, red), and after MVA boost (week 50, green). Coloured squares indicate positive responses. 'x' indicates missing data as a result of insufficient PBMCs.



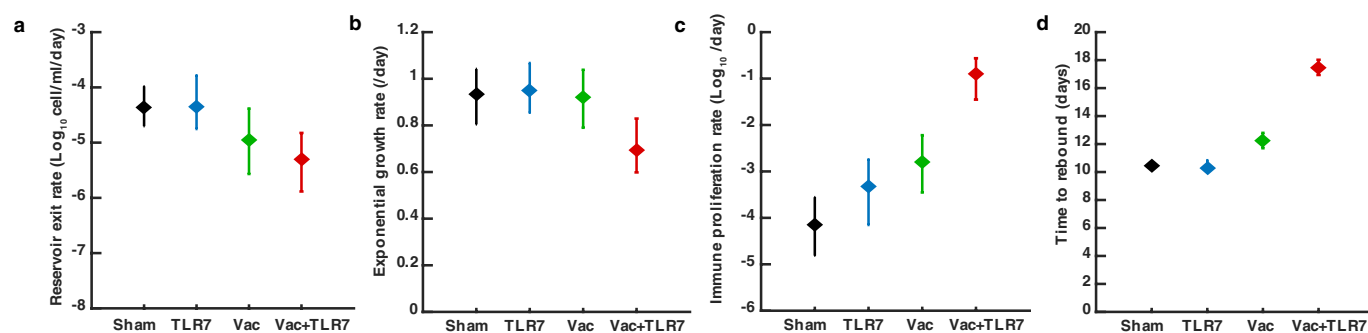
Extended Data Figure 7 | Correlations of cellular immune breadth to day 7 SIV RNA. Correlations are shown for the breadth of Gag, Pol, Env and total (Gag+Pol+Env) cellular immune responses at the time of ART discontinuation at week 72 and pre-ART day 7 log SIV RNA. *P* value indicates two-sided Spearman rank-correlation test. *R* value indicates correlation coefficient.



Extended Data Figure 8 | Correlations of cellular immune breadth to set point viral loads. Correlations are shown for the breadth of Gag, Pol, Env and total (Gag+Pol+Env) cellular immune responses at peak immunity at week 50 and set point log SIV RNA following ART discontinuation. *P* values indicate two-sided Spearman rank-correlation tests. *R* values indicate correlation coefficients.



Extended Data Figure 9 | Correlations of cellular immune breadth to time to viral rebound. Correlations are shown for the breadth of Gag, Pol, Env and total (Gag+Pol+Env) cellular immune responses at peak immunity at week 50 and time to viral rebound following ART discontinuation. P values indicate two-sided Spearman rank-correlation tests. R values indicate correlation coefficients.



Extended Data Figure 10 | Rebound kinetic parameters estimated from viral dynamics modelling. Viral load values following ART discontinuation in each animal were fit to a viral dynamics model using a Bayesian framework. **a–d**, Plots show the median and 95% credible intervals for estimations of the rate of reactivation of cells from the latent reservoir (**a**), the initial exponential growth rate (**b**), the immune proliferation rate (**c**), and the time at which viral load reaches the detection

threshold of 200 copies per ml for each treatment group (**d**). Monkeys treated with both the vaccine and TLR7 agonist exhibited slower viral growth rates and stronger immune responses than all other groups ($P < 0.01$ for each comparison). These monkeys and monkeys treated with only the vaccine exhibited a lower reservoir exit rate than untreated monkeys ($P < 0.05$ for each comparison).

Host genome integration and giant virus–induced reactivation of the virophage mavirus

Matthias G. Fischer¹ & Thomas Hackl^{1†}

Endogenous viral elements are increasingly found in eukaryotic genomes¹, yet little is known about their origins, dynamics, or function. Here we provide a compelling example of a DNA virus that readily integrates into a eukaryotic genome where it acts as an inducible antiviral defence system. We found that the virophage mavirus², a parasite of the giant *Cafeteria roenbergensis* virus (CroV)³, integrates at multiple sites within the nuclear genome of the marine protozoan *Cafeteria roenbergensis*⁴. The endogenous mavirus is structurally and genetically similar to eukaryotic DNA transposons and endogenous viruses of the Maverick/Polinton family^{5–7}. Provirophage genes are not constitutively expressed, but are specifically activated by superinfection with CroV, which induces the production of infectious mavirus particles. Virophages can inhibit the replication of mimivirus-like giant viruses and an anti-viral protective effect of provirophages on their hosts has been hypothesized^{2,8}. We find that provirophage-carrying cells are not directly protected from CroV; however, lysis of these cells releases infectious mavirus particles that are then able to suppress CroV replication and enhance host survival during subsequent rounds of infection. The microbial host–parasite interaction described here involves an altruistic aspect and suggests that giant-virus-induced activation of provirophages might be ecologically relevant in natural protist populations.

Virophages of the family *Lavidaviridae*⁹ are obligate parasites of giant DNA viruses of the family *Mimiviridae*. During co-infection of a protist host, these 15–30 kilobase pair (kbp) double-stranded DNA (dsDNA) viruses replicate in the cytoplasmic virion factory of the giant virus, which can result in decreased giant viral progeny and increased host survival rates^{2,10}. Mavirus-like virophages share an evolutionary origin with a class of self-synthesizing DNA transposons called Maverick/Polinton elements^{2,5,6}; however, the nature of the common ancestor is a matter of debate^{2,7}. Maverick/Polinton elements are widespread in eukaryotic genomes and encode virus-like genes, which led to their recent designation as ‘polintoviruses’, representing perhaps the most broadly distributed family of endogenous viral elements among eukaryotes^{7,11,12}. Virophages encode integrase genes, and endogenous genomes (‘provirophages’) have been reported in mimiviruses^{13,14} and in the nuclear genome of the marine alga *Bigelowiella natans*⁸. Provirophages in protists are hypothesized to protect the host cells from giant-virus infection^{2,8}, but experimental data for this claim were lacking.

To generate host-integrated provirophages, we infected the clonal *C. roenbergensis* strain E4-10P with the giant virus CroV at a multiplicity of infection (MOI) of 0.01 and with mavirus at MOI ≈ 1 and screened the surviving cells for mavirus DNA (Fig. 1a and Extended Data Fig. 1). PCR analysis of 66 clonal survivor strains identified 21 (32%) mavirus-positive host strains. We chose the cell line with the strongest mavirus signal for further analysis and named it E4-10M1. Filtration of E4-10P and E4-10M1 cell populations through syringe filters of various pore sizes and quantitative PCR (qPCR) analysis of

the resulting filtrates confirmed that the observed mavirus signal was associated with E4-10M1 cells and not caused by remaining free virus particles (Extended Data Table 1). We then sequenced genomic DNA from strains E4-10P and E4-10M1 on Illumina MiSeq and Pacific BioSciences (PacBio) RS II platforms and created hybrid assemblies for each strain. The sequence data suggested that *C. roenbergensis* has a diploid genome (Extended Data Fig. 2), which obstructed the direct assembly of mavirus-containing contigs because integration at a specific site occurred at only one of the two alleles, thus introducing a structural ambiguity that led to erroneous assemblies. We therefore scanned the E4-10M1 genome assembly indirectly for integrated mavirus sequences by aligning corrected PacBio reads to the mavirus reference genome, extracting those reads, and assembling them into contigs. The longest resulting contig was 30,556 bp in length and contained a 19,055-bp sequence that was 100% identical to the mavirus reference genome (GenBank accession number HQ712116). In contrast to the reference mavirus genome, the endogenous virus genome was flanked on either side by 615/616-bp-long terminal inverted repeats (TIRs) that were 99.7% identical to each other. The longer TIRs result in a total length of 20,190 bp for the endogenous mavirus genome, compared with 19,063 bp for the reference genome. By recruiting reads to the flanking regions of mavirus TIRs, we found 11 well-supported integration sites in the E4-10M1 genome (Extended Data Table 2). The host sequence directly adjacent to the provirophage genome featured target site duplications that were 5–6 bp long. The target site duplication sequences differed between integration sites with no obvious consensus motif. One of the integration sites was characterized in further detail and its reconstruction is shown in Fig. 1b. PCR analysis verified the predicted integration site and confirmed that the E4-10M1 strain is heterozygous for the integrated mavirus genome (Fig. 1c, d).

To test whether the endogenous mavirus genes were expressed, we analysed selected transcripts by reverse transcription (RT) qPCR. Mavirus gene promoters are highly similar to the late gene promoter motif in CroV², which suggests that CroV can activate mavirus genes. We therefore isolated total RNA from mock-infected and CroV-infected E4-10P and E4-10M1 cells at 0 and 24 h post infection (p.i.) and quantified in the resulting complementary DNA (cDNA) pool five mavirus genes, three CroV genes, and one host cell gene using gene-specific primers (Extended Data Table 3). CroV transcripts could be clearly detected at 24 h p.i. in the infected cultures and their expression levels were comparable between E4-10P and E4-10M1 strains (Fig. 2). The mavirus genes in E4-10M1 cells were quiescent under normal conditions and immediately after inoculation with CroV. However, mavirus genes were strongly expressed at 24 h p.i. in the CroV-infected E4-10M1 strain. CroV infection thus induces expression of the endogenous mavirus genes in E4-10M1 cells. Addition of the protein biosynthesis inhibitor cycloheximide (CHX) or the DNA polymerase inhibitor aphidicolin effectively inhibited host cell growth and CroV DNA replication (Supplementary Spreadsheet 1). CHX treatment inhibited expression of the intermediate DNA polymerase B gene *croV497* and of

¹Department of Biomolecular Mechanisms, Max Planck Institute for Medical Research, 69120 Heidelberg, Germany. [†]Present address: Massachusetts Institute of Technology, 15 Vassar Street, Cambridge, Massachusetts 02139, USA.

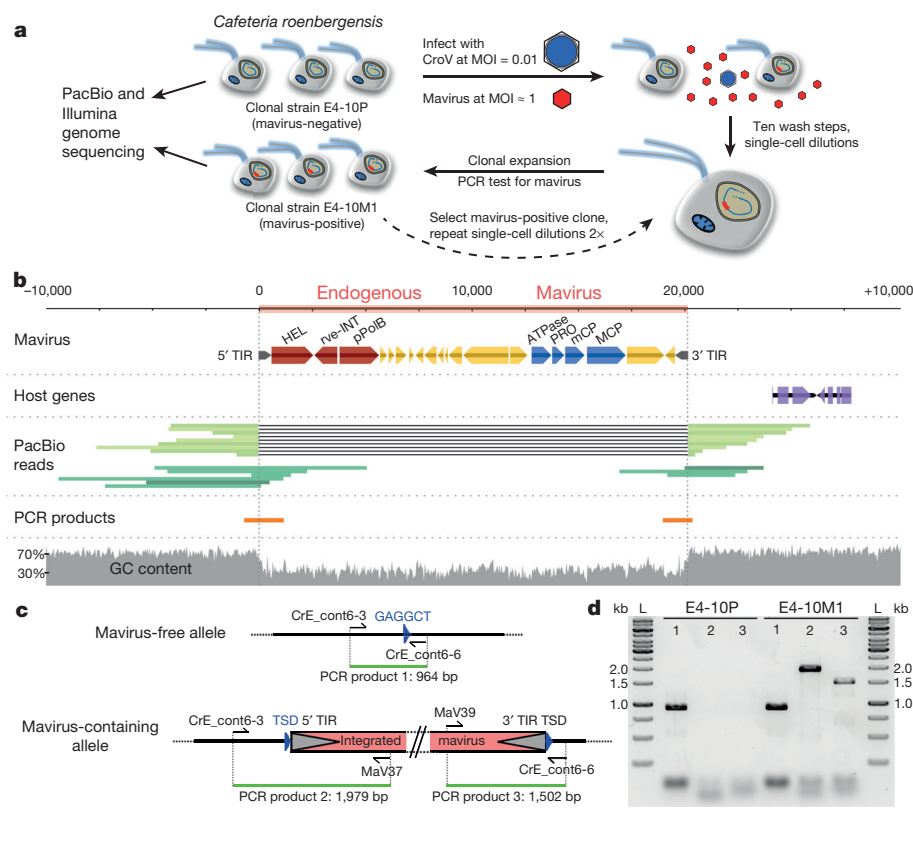


Figure 1 | Generation and characterization of endogenous mavirus virophages in *Cafeteria roenbergensis*. **a**, The clonal *C. roenbergensis* strain E4-10P was infected with CroV and mavirus and surviving cells were PCR-tested for mavirus. The genomes of the resulting mavirus-positive clonal strain E4-10M1 and of the parental E4-10P strain were sequenced on PacBio and Illumina platforms to analyse mavirus integration sites. **b**, Partial view of a 208-kbp E4-10M1 contig featuring the 20,190-bp mavirus genome flanked by 10 kbp of host sequence on either side. Mavirus genes for replication and integration are shown in red, morphogenetic genes are shown in blue, other genes are shown in yellow, and TIRs are indicated in grey. The exon structure of two adjacent host gene models (function unknown) is shown in purple. PacBio reads covering the integration site are shown in green. Reads that span the integration site and contain only host sequence are shown in light green, whereas reads that cross the virus–host junction are shown in dark green. PCR products spanning the host–virus junction are shown in orange. The GC content plot is based on a 30-bp sliding window. **c**, Representation of the host genomic region in **b** illustrating the PCR primer binding sites and expected PCR products that were used to confirm the integration site. **d**, Gel image of the PCR products obtained from E4-10P and E4-10M1 genomic templates with primers spanning the integration site. The lanes are labelled according to the primer combinations and products shown in **c**. L, DNA ladder.

the late major capsid gene *croV342*, but not of the early isoleucyl-tRNA synthetase gene *croV505* (Extended Data Fig. 3). This is in line with the presence of a viral transcription apparatus in the virion of CroV¹⁵ and other cytoplasmic large DNA viruses^{16–18}, which mediates early viral gene expression. In the presence of aphidicolin, all three CroV genes were expressed at low levels, with cDNA of the late MCP gene

being barely detectable. Notably, treatment with CHX or aphidicolin also inhibited mavirus gene expression in CroV-infected E4-10M1 cells, indicating that *de novo* protein synthesis and CroV DNA replication are prerequisites for proviophage gene induction. On the basis of these results and the similarity of transcriptional signals between virophages and viruses of the *Mimiviridae* family,

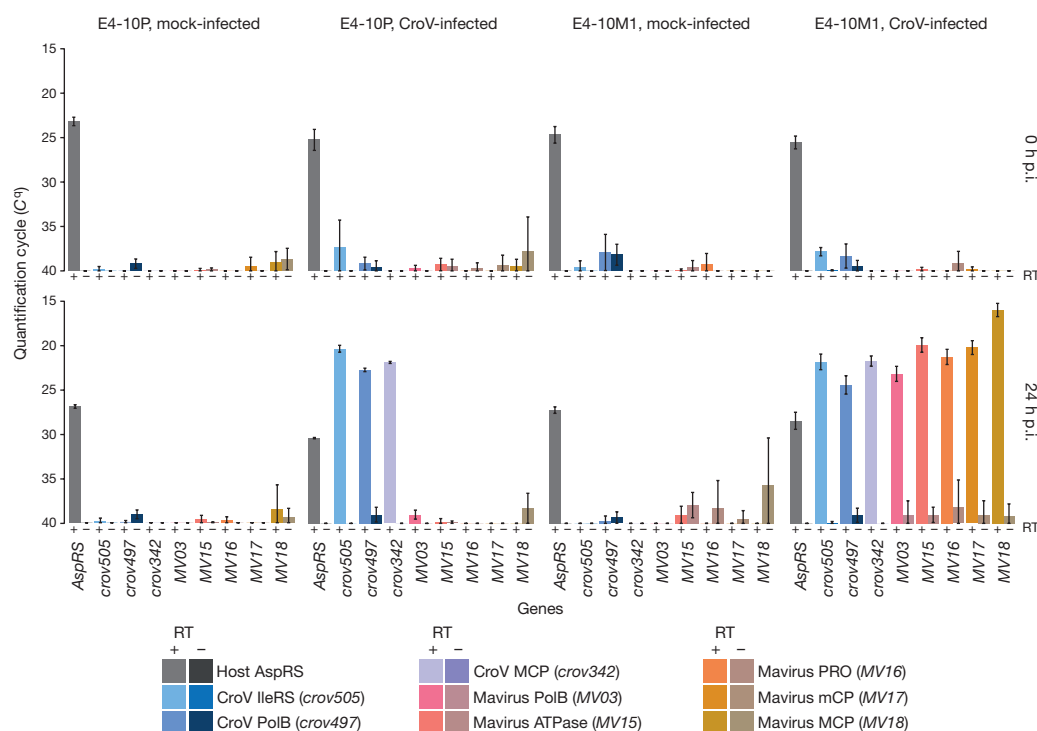


Figure 2 | Gene expression analysis of the endogenous mavirus genome. Selected cellular and viral transcripts isolated at 0 and 24 h p.i. from mock-infected or CroV-infected E4-10P and E4-10M1 cultures were quantified by qRT-PCR. Shown are the average quantification cycle (C_q) values of three independent experiments; error bars, s.d. The following genes were assayed: host AspRS, *C. roenbergensis* E4-10 aspartyl-tRNA synthetase; *croV342*, CroV major capsid protein; *croV497*, CroV DNA polymerase B; *croV505*, CroV isoleucyl-tRNA synthetase; MV03, mavirus DNA polymerase B; MV15, mavirus genome-packaging ATPase; MV16, mavirus maturation protease; MV17, mavirus minor capsid protein; MV18, mavirus major capsid protein. C_q values of the reverse transcriptase (RT) negative reactions are shown directly to the right of the respective RT-positive results. Accession numbers are listed in Extended Data Table 3. See also Supplementary Spreadsheet 1.

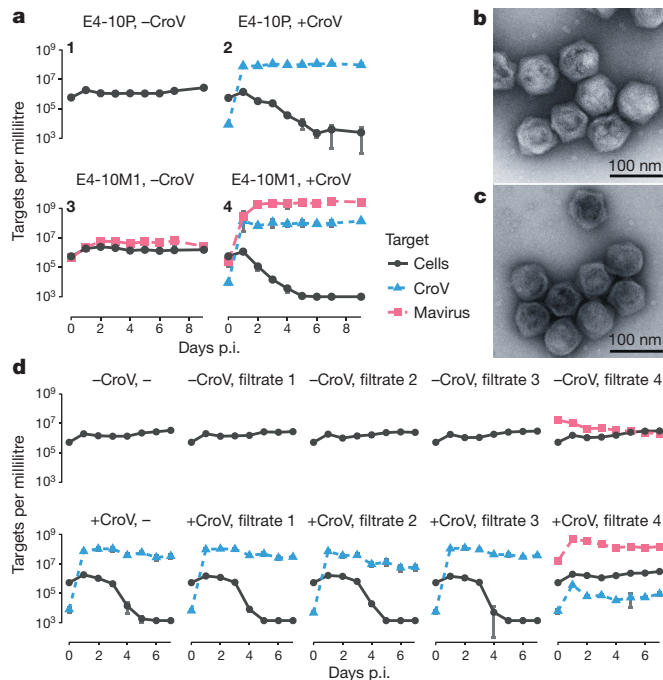


Figure 3 | CroV infection induces replication and virion production of the endogenous mavirus. **a**, *C. roenbergensis* strains E4-10P and E4-10M1 were mock-infected (infections 1, 3) or infected with CroV (infections 2, 4) and cells and viruses were monitored for 9 days. Cell densities are based on microscopy counts; virus concentrations are derived from qPCR data. **b**, Negative-stain electron micrograph of virus particles from the CroV-infected E4-10M1 strain. **c**, Electron micrograph of reference mavirus particles. **d**, Strain E4-10P was mock-infected (upper row) or CroV-infected (lower row) and simultaneously inoculated with 0.02% (v/v) of 0.1 µm pore-size filtered samples taken at 3 d p.i. during the infection experiments 1–4 shown in **a**. Data in **a** and **d** were pooled from three independent experiments; error bars, s.d. See also Supplementary Spreadsheet 2.

we hypothesize that a CroV-encoded late transcription factor might be responsible for proviophage activation (Extended Data Fig. 4), although we do not exclude alternative pathways.

Next, we examined whether CroV infection would induce DNA replication of the integrated mavirus genome. E4-10P and E4-10M1 cells were either mock-infected or CroV-infected and viral DNA levels were monitored by qPCR. No virus DNA was found in mock-infected E4-10P cells, whereas a latent mavirus signal was present in mock-infected E4-10M1 cells (Fig. 3a). In CroV-infected E4-10M1 cells, the mavirus signal increased approximately 500-fold within 48 h p.i., proving that CroV induces genome replication of the mavirus proviophages. CroV replication and cell lysis were comparable in both host strains and the CroV titre was similar in E4-10P and E4-10M1 lysates (about $5 \times 10^7 \text{ ml}^{-1}$), indicating that proviophage induction does not inhibit CroV propagation or prevent cell lysis in CroV-infected E4-10M1 cells. Electron microscopy of concentrated and purified cell lysates revealed the presence of mavirus-like particles in CroV-infected E4-10M1, but not E4-10P lysates, nor in mock-infected cultures (Fig. 3b, c and Extended Data Fig. 5). To test whether these particles are infectious, we co-inoculated E4-10P cultures with 0.1-µm-filtered material from mock- or CroV-infected E4-10P and E4-10M1 cultures (Fig. 3d). Only lysate from CroV-infected E4-10M1 cultures contained mavirus DNA, which replicated in the presence of CroV. We conclude from these results that CroV induces the production of infectious mavirus particles in strain E4-10M1. Interestingly, the reactivated mavirus suppressed CroV genome replication by two to three orders of magnitude, resulting in survival of the host cell population (Fig. 3d). Treatment of reactivated mavirus with 500 J m^{-2} of ultraviolet light ($\lambda = 254 \text{ nm}$) before infection abrogated these effects (Extended Data Fig. 6), suggesting that mavirus is indeed the causative agent.

To gain more insight in the virophage–virus–host dynamics, we infected E4-10P cells with different MOIs of CroV and of reactivated mavirus. Figure 4a shows infections with CroV at MOIs of 0.01–10 in the absence or presence of mavirus at MOI ≈ 10 . The number of virions that each cell receives at a given MOI follows a Poisson distribution; therefore the percentage of infected cells at an MOI of 1 is 63%, and an MOI of 10 is needed to ensure that more than 99.9% of cells

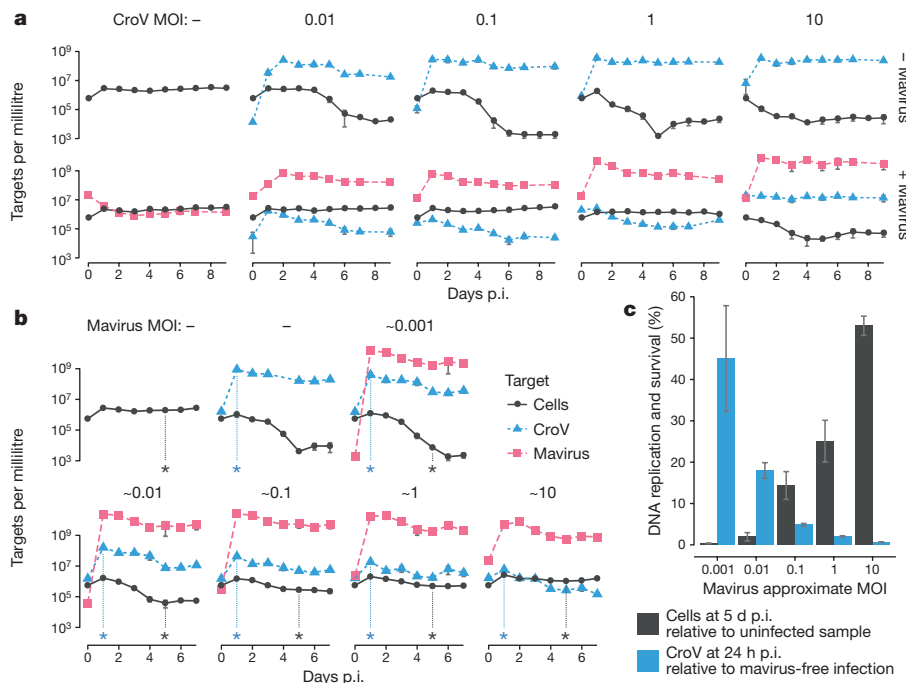


Figure 4 | Reactivated mavirus inhibits CroV and promotes host survival in subsequent infections. *C. roenbergensis* strain E4-10P

were infected with different MOIs of CroV and mavirus. Mavirus inocula represent the 0.1 µm filtrate of the CroV-infected E4-10M1 culture at 3 d p.i. (Fig. 3a, infection 4). Cell densities are based on microscopy counts; virus concentrations are derived from qPCR data. Shown are the average values of biological triplicates; error bars, s.d. See also Supplementary Spreadsheet 3.

a, Infection experiments with different MOIs of CroV in the absence (upper row) or presence (lower row) of mavirus at an MOI of about 10.

b, Cultures of *C. roenbergensis* strain E4-10P infected with CroV at MOI = 1 and increasing MOIs of reactivated mavirus. The first panel shows the mock-infected control; the second panel shows the mavirus-free, CroV-infected control. Asterisks and vertical dotted lines mark the reference points for the analysis in **c**. **c**, Summary of the effect of mavirus MOI on CroV DNA replication and host cell survival from the infection experiments shown in **b**. Black columns show the host cell densities at 5 d p.i. with increasing mavirus MOIs relative to mock-infected 5 d p.i. cultures. Blue columns show the CroV genome copy concentration at 24 h p.i. with increasing mavirus MOIs relative to mavirus-free 24 h p.i. CroV infections.

are infected. With every cell infected with mavirus, host populations survived an infection with CroV at MOIs of 0.01–1 (Fig. 4a). Although CroV at MOI = 10 did not replicate in the presence of mavirus, the cells still lysed (96.5% decline after 5 days). These data indicate that nearly every cell infected with CroV dies, irrespective of mavirus, and that mavirus rather halts the spread of CroV by inhibiting its replication and preventing the release of progeny virions from lysed co-infected cells. When E4-10P cells were infected with CroV at MOI = 1 and mavirus at MOIs ranging from approximately 0.001 to 10, a clear dose–response relationship was observed for host survival and inhibition of CroV DNA replication (Fig. 4b). Mavirus inhibited CroV and improved host survival rates even at low MOIs (Fig. 4c).

Our results show that mavirus readily integrates in the nuclear genome of its eukaryotic host *C. roenbergensis* and that mavirus proviophages resemble Maverick/Polinton elements not only in length, gene content, and host range, but also in their DNA repeats. We demonstrate that endogenous mavirus genes are transcriptionally silent unless the cell is infected with CroV, which triggers gene expression, genome replication, and virion synthesis from the proviophage. Although proviophage-carrying cells are not directly protected from CroV, lysis of these cells releases infectious mavirus particles that are able to inhibit CroV in subsequent co-infections. Proviophage-mediated host defence against giant viruses in the *Cafeteria* system thus appears to follow an altruistic model, in which some cells are sacrificed to protect their kin. Our study also reveals a mutualistic virus–host relationship, in which the cell provides an opportunity for mavirus to persist as a proviophage while the host population may benefit from mavirus in the presence of CroV. It remains to be tested how widespread proviophages are in natural *Cafeteria* populations and to what extent they influence the ecology of this important group of marine grazers.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 June; accepted 2 November 2016.

1. Aiewsakun, P. & Katourakis, A. Endogenous viruses: connecting recent and ancient viral evolution. *Virology* **479–480**, 26–37 (2015).
2. Fischer, M. G. & Suttle, C. A. A virophage at the origin of large DNA transposons. *Science* **332**, 231–234 (2011).
3. Fischer, M. G., Allen, M. J., Wilson, W. H. & Suttle, C. A. Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc. Natl Acad. Sci. USA* **107**, 19508–19513 (2010).
4. Fenchel, T. & Patterson, D. J. *Cafeteria roenbergensis* nov. gen., nov. sp., a heterotrophic microflagellate from marine plankton. *Mar. Microb. Food Webs* **3**, 9–19 (1988).
5. Pritham, E. J., Putliwala, T. & Feschotte, C. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* **390**, 3–17 (2007).

6. Kapitonov, V. V. & Jurka, J. Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl Acad. Sci. USA* **103**, 4540–4545 (2006).
7. Krupovic, M. & Koonin, E. V. Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. *Nature Rev. Microbiol.* **13**, 105–115 (2015).
8. Blanc, G., Gallot-Lavallée, L. & Maumus, F. Proviophages in the *Bigelowiella* genome bear testimony to past encounters with giant viruses. *Proc. Natl Acad. Sci. USA* **112**, E5318–E5326 (2015).
9. Krupovic, M., Kuhn, J. H. & Fischer, M. G. A classification system for virophages and satellite viruses. *Arch. Virol.* **161**, 233–247 (2016).
10. La Scola, B. et al. The virophage as a unique parasite of the giant mimivirus. *Nature* **455**, 100–104 (2008).
11. Krupovic, M. & Koonin, E. V. Self-synthesizing transposons: unexpected key players in the evolution of viruses and defense systems. *Curr. Opin. Microbiol.* **31**, 25–33 (2016).
12. Yutin, N., Shevchenko, S., Kapitonov, V., Krupovic, M. & Koonin, E. V. A novel group of diverse Polinton-like viruses discovered by metagenome analysis. *BMC Biol.* **13**, 95 (2015).
13. Desnues, C. et al. Proviophages and transpovirons as the diverse mobilome of giant viruses. *Proc. Natl Acad. Sci. USA* **109**, 18078–18083 (2012).
14. Santini, S. et al. Genome of *Phaeocystis globosa* virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. *Proc. Natl Acad. Sci. USA* **110**, 10800–10805 (2013).
15. Fischer, M. G., Kelly, I., Foster, L. J. & Suttle, C. A. The virion of *Cafeteria roenbergensis* virus (CroV) contains a complex suite of proteins for transcription and DNA repair. *Virology* **466–467**, 82–94 (2014).
16. Resch, W., Hixson, K. K., Moore, R. J., Lipton, M. S. & Moss, B. Protein composition of the vaccinia virus mature virion. *Virology* **358**, 233–247 (2007).
17. Renesto, P. et al. Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. *J. Virol.* **80**, 11678–11685 (2006).
18. Legendre, M. et al. In-depth study of *Mollivirus sibericum*, a new 30,000-y-old giant virus infecting *Acanthamoeba*. *Proc. Natl Acad. Sci. USA* **112**, E5327–E5335 (2015).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by the Max Planck Society. We are grateful to C. Suttle for access to host and virus strains, and to the Roscoff team for maintaining and distributing protist strains. We thank K.-A. Seifert, K. Fenzl and K. Barenhoff for technical assistance, U. Mersdorf for electron microscopy, C. Roome for IT support, L. Czaja and the Max Planck Genome Centre in Cologne for bioinformatic assistance, S. Higgins for suggestions, K. Haslinger and J. Reinstein for comments on the manuscript, and I. Schlichting for mentoring and support.

Author Contributions M.G.F. conceived the study, designed and performed experiments, collected, analysed, and interpreted data, and wrote the manuscript. T.H. corrected and assembled sequence data, and analysed, interpreted, and visualized data. Both authors revised and approved the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.G.F. (mfischer@mpimf-heidelberg.mpg.de).

Reviewer Information *Nature* thanks J.-M. Claverie, E. Koonin and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Host and virus strains. *C. roenbergensis* strain E4-10 was isolated from coastal waters near Yaquina Bay, Oregon, USA, as described previously¹⁹. The cell suspension culture has since been continuously passaged approximately every 4 weeks in f/2 enriched natural or artificial seawater medium supplemented with one to three autoclaved wheat grains per 10 ml to stimulate bacterial growth. For infection experiments, cells were grown in f/2 enriched artificial seawater medium supplemented with 0.05% (w/v) Bacto yeast extract (Becton, Dickinson, Germany). For f/2 artificial seawater medium, the following sterile stock solutions were prepared: 75 g/l NaNO₃; 5 g/l NaH₂PO₄; 1,000× trace metal solution containing 4.36 g/l Na₂EDTA × 2 H₂O, 3.15 g/l FeCl₃ × 6 H₂O, 0.01 g/l CuSO₄ × 5 H₂O, 0.18 g/l MnCl₂ × 4 H₂O, 0.006 g/l Na₂MoO₄ × 2 H₂O, 0.022 g/l ZnSO₄ × 7 H₂O, 0.01 g/l CoCl₂ × 6 H₂O; and a 50,000× vitamin solution containing 5 g/l thiamine-HCl, 25 mg/l biotin, and 25 mg/l cyanocobalamin. The vitamin solution was stored at −20°C, the other solutions at room temperature. To prepare 1 l of f/2 artificial seawater medium, 33 g of Red Sea Salt (Red Sea Meersalz, www.aquaristikshop.com) were dissolved in ultra-pure water (ELGA, Veolia Water Technologies, Germany), then 1 ml each of the 75 g/l NaNO₃, 5 g/l NaH₂PO₄, and 1,000× trace metal solutions as well as 20 µl of the 50,000× vitamin solution were added. After autoclaving, the medium was 0.22-µm filtered and stored at 4°C. Cultures were grown in flat-bottom 125 ml or 250 ml polycarbonate Erlenmeyer flasks (VWR, Germany) at 23°C in the dark.

The viruses used for infection experiments were Cafeteria roenbergensis virus (CroV) strain BV-PW1 (refs 3, 20) and mavirus strain Spezl².

Viral infectivity assays. The infectivity of CroV was measured by end-point dilution assays and the statistical method by Reed and Muench²¹ was used to determine the 50% end point. The resulting cell culture infectious dose at which 50% of the cultures lysed (CCID₅₀) was within a factor of four of counts of SYBR-stained CroV particles by epifluorescent microscopy and within a factor of two of gene copy numbers derived by qPCR. End-point dilution assays were performed in 96-well plates with 200 µl of 1 × 10⁶ cells per millilitre exponentially growing host cells in f/2 medium + 0.05% (w/v) yeast extract per well. Each row (12 wells) was inoculated with a different dilution of CroV suspension (10 µl per well). Dilutions ranged from 10^{−2} to 10^{−9}. The plates were stored at 23°C and analysed after 6 days for cell lysis by microscopy. For mavirus, end-point dilution assays could not be employed because, in contrast to CroV, a productive mavirus infection does not result in cell lysis or cytopathic effects.

Infection experiments. Typically, host cell suspension cultures were diluted daily to a cell density of 1 × 10⁵ to 5 × 10⁵ cells per millilitre with f/2 medium containing 0.05% (w/v) yeast extract, until the desired culture volumes were reached. On the day of infection, when the cells had reached a density of >10⁶ cells per millilitre, the cultures were diluted with f/2 medium containing 0.05% (w/v) yeast extract to a cell density of 5 × 10⁵ to 7 × 10⁵ cells per millilitre. Depending on the experiment, aliquots of 20 ml or 50 ml were dispensed in 125 ml or 250 ml polycarbonate flat-base Erlenmeyer flasks (Corning, Germany; through VWR International) and inoculated with virus-containing lysate or virus-free f/2 medium (for mock infections). The CroV inoculum varied between different infection experiments according to the desired MOI and the titre of the CroV working stock, which was stored at 4°C and replaced every few months. Mock-infected cultures received an equal volume of f/2 medium. For testing culture supernatant from previous infection experiments for mavirus activity, 1 ml of the appropriate 0.1 µm pore-size filtered lysate was added to the flask immediately before the CroV inoculum. Cultures were incubated at 23°C in the dark. Cell concentrations were measured by staining a 10 µl aliquot of the suspension culture with 1 µl of Lugol's acid iodine solution and counting the cells on a haemocytometer (Neubauer Improved Counting Chamber, VWR Germany). This method does not distinguish between live and dead cells and includes cells that are already dead but have not lysed yet. Aliquots (200 µl) for DNA extraction were sampled at appropriate time points and were immediately frozen and stored at −20°C until further processing. All infections were performed in triplicate, except those shown in Extended Data Fig. 1, which were done in single replicates.

Isolation of *C. roenbergensis* strains E4-10P and E4-10M1. *C. roenbergensis* strain E4-10 was made clonal by repeated single-cell dilutions. Each well of a 96-well plate was filled with 200 µl of f/2 medium containing 0.01% (w/v) yeast extract. Then 1 µl of an E4-10 culture diluted to 300 cells per millilitre was added to each well, so that on average every third well received one cell. After 6 days at 23°C, wells were inspected for cell growth and positive samples were transferred to 20 ml of f/2 medium containing 0.05% (w/v) yeast extract. This procedure was repeated serially twice more. DNA from the final isolate was extracted and tested by qPCR to confirm the absence of mavirus. Cultures (20 ml) of the resulting E4-10P

(parental) strain at 5 × 10⁵ cells per millilitre in f/2 medium containing 0.05% (w/v) yeast extract were then either mock-infected, infected with CroV at MOI = 0.01, or co-infected with CroV (MOI = 0.01) and mavirus (MOI ≈ 1). Eight days after infection, the surviving cells from the co-infection were pelleted by centrifugation (5 min at 7,000g, 23°C), the pellets were resuspended in 50 ml f/2 medium, and the centrifugation/dilution procedure was repeated nine more times. The washed cells were then subjected to three consecutive rounds of single-cell dilution as described above. DNA was extracted from the resulting 66 clonal strains and tested by qPCR with mavirus-specific primers. The strain with the highest qPCR signal was named E4-10M1 (first mavirus-positive strain).

Filtration assay. Host strains E4-10P and E4-10M1 were either infected with CroV or mock-infected with f/2 medium. At 5 days p.i., when the CroV-infected cells had lysed, aliquots from the four different samples were passed through syringe filters of different nominal pore sizes, ranging from 5.0 µm to 0.1 µm, and DNA was extracted from 200 µl of each filtrate as well as from 200 µl of the unfiltered samples. The following syringe filters were used: 0.1 µm pore-size PVDF Millex (Millipore Merck, Ireland), 0.22 µm pore-size PES (TPP, Switzerland), 0.45 µm pore-size PES (TPP, Switzerland), 5.0 µm pore-size CN-S Whatman (Fisher Scientific, Germany). E4-10M1 cells were mechanically lysed by sonication with a Branson Sonifier 250 equipped with a microtip, duty cycle 50%, output setting 2. Two millilitre aliquots of an E4-10M1 suspension culture containing 1.4 × 10⁶ cells per millilitre were sonicated twice for 30 s with incubation on ice for 30 s in between. As a positive control, an E4-10P suspension culture was mixed with 0.1-µm-filtered reactivated mavirus to yield a final flagellate concentration of 1.8 × 10⁶ cells per millilitre. The sonicated and positive control samples were then filtered and processed as described above.

DNA extraction and qPCR. We used qPCR with the SYBR-related EvaGreen dye to quantify viral DNA target sequences. Genomic DNA (gDNA) was extracted from 200 µl of suspension culture with the DNeasy 96 Blood & Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions for DNA purification of total DNA from cultured cells, with a single elution step in 100 µl of double-distilled (dd) H₂O and storage at −20°C. DNA concentrations in the eluted samples typically ranged from 1 to 10 ng/µl, as measured on a NanoDrop 2000c spectrophotometer (Thermo Scientific, Germany). The variations are largely due to fluctuations in the bacterial populations that were present in the flagellate cultures. One microlitre of gDNA was used as template in a 20-µl qPCR reaction containing 10 µl of 2 × Fast-Plus EvaGreen Master Mix with low ROX dye (Biotium, Inc. via VWR, Germany), 10 pmol of each forward and reverse primer (see Supplementary Table 3), and 8.8 µl of ddH₂O. No-template controls (NTC) contained ddH₂O instead of gDNA. Each qPCR reaction (sample, NTC, or standard) was performed in technical duplicates, with individual replicates differing in their quantification cycles (C_q) by about 0.5% on average (0.49% ± 0.43%, *n* = 200). The limit of detection for this assay was ≈10 copies, which equates to about 5,000 copies per millilitre of suspension culture. The C_q values of the NTC controls were consistently below the limit of detection. Thermal cycling was performed in a Stratagene Mx3005P qPCR system (Agilent Technologies, Germany) with the following settings: 95°C for 5 min, 40 cycles of 95°C for 10 s followed by 60°C for 25 s and 72°C for 25 s, a single cycle of 72°C for 5 min, and a final dissociation curve was recorded from 50°C to 95°C. qPCR results were analysed using MxPro qPCR software version 4.10 (Stratagene, La Jolla, California, USA). The threshold fluorescence was set using the amplification-based option of MxPro software. Standard curves were calculated from a tenfold dilution series that ranged from 10¹ to 10⁸ molecules of a linearized pEX-A plasmid (Eurofins Genomics, Germany) carrying the fragment of the MV18 MCP gene (GenBank accession number ADZ16417) that was amplified by primers Spezl-qPCR-5 and Spezl-qPCR-6 (Extended Data Table 3) for mavirus quantification, or gDNA extracted from a known amount of CroV particles, the concentration of which had been determined by epifluorescence microscopy. To compare the two different kinds of template DNA used for virus quantification directly, the linearized plasmid also contained the target sequence for the *crov283* gene (GenBank accession number ADO67316.1) that was amplified by primers CroV-qPCR-9 and CroV-qPCR-10 and used as an approximation for CroV genome copies. The resulting standard curves and C_q values of the plasmid and gDNA templates were nearly identical, which implies that the quantification of mavirus using a plasmid-encoded target sequence is a valid approach. For mavirus quantification with primers Spezl-qPCR-5 and Spezl-qPCR-6, the R² value for the standard curve was 0.996, the amplification efficiency was 109.7%, and the standard curve equation was $y = -3.109\log(x) + 33.89$. For CroV quantification with primers CroV-qPCR-9 and CroV-qPCR-10, the R² value for the standard curve was 1.000, the amplification efficiency was 103.0%, and the standard curve equation was $y = -3.253\log(x) + 34.77$.

PCR verification of an example mavirus integration site. The mavirus integration site shown in Fig. 1b was verified by PCR analysis and Sanger sequencing of the PCR products. Owing to the difficulty of obtaining PCR products that were part

host sequence with 70% GC content and part mavirus sequence with 30% GC content, several primers had to be tested under various PCR cycling conditions before the predicted products could be obtained. Primer sequences are listed in Extended Data Table 3. PCR amplifications were performed using 2 ng of genomic DNA template from strain E4-10P or E4-10M1 in a 25 µl reaction mix containing 5 µl Q5 Reaction Buffer (NEB, Germany), 0.5 U of Q5 High-Fidelity DNA Polymerase (NEB, Germany), 0.2 mM dNTPs, and 0.5 µM of each primer. In addition, the PCR mixes to amplify the empty integration site with primers CrE_cont6-3 and CrE_cont6-6 (amplifying only host sequence with 70% GC content) contained 5 µl of Q5 High GC Enhancer solution. The PCRs were performed in a TGradient thermocycler (Biometra, Germany) with the following cycling conditions: 30 s denaturation at 98 °C; 35 cycles of 10 s denaturation at 98 °C, 30 s annealing at 68 °C (for primer pair CrE_cont6-3 & MaV37) or 69 °C (for primer pairs MaV39 & CrE_cont6-6 and CrE_cont6-3 & CrE_cont6-6) and 1 min extension at 72 °C; and a final 2 min extension at 72 °C. For product analysis, 5 µl of each reaction were mixed with loading dye and pipetted on a 1% (w/v) agarose gel supplemented with GelRed. The marker lanes contained 0.5 µg of GeneRuler 1 kb DNA Ladder (Fermentas, Thermo-Fisher Scientific, USA). The gel was electrophoresed for 2 h at 70 V and visualized on a ChemiDoc MP Imaging System (BioRad, Germany).

Cycling conditions for the PCR shown in Extended Data Figure 5 were 45 s denaturation at 98 °C; 35 cycles of 10 s denaturation at 98 °C, 30 s annealing at 58 °C (primer pairs MaV21F and MaV21R) and 1 min extension at 72 °C; and a final 2 min extension at 72 °C.

RNA extraction and qRT-PCR. Triplicate 50 ml cultures of strains E4-10P and E4-10M1 at an initial cell density of 6×10^5 cells per millilitre were either mock-infected with f/2 medium or infected with CroV at MOI = 0.2. Aphidicolin-treated cultures were supplemented with 125 µl of a 2 mg/ml aphidicolin solution in DMSO (Sigma-Aldrich, Germany) for a final concentration of 5 µg/ml. Cycloheximide-treated cultures were supplemented with 37.5 µl of a 66.6 mg/ml CHX solution in DMSO (Sigma-Aldrich, Germany) for a final concentration of 50 µg/ml. Cultures were incubated at 23 °C. For extraction of total RNA, 1 ml aliquots were taken from each culture at 0 h p.i. and 24 h p.i. and centrifuged for 5 min at 10,000g, 21 °C. The supernatants were discarded and the cell pellets were immediately flash-frozen in N₂(l) and stored at -80 °C until further use. RNA extraction was performed with the Qiagen RNeasy Mini Kit following the protocol for purification of total RNA from animal cells using spin technology. Cells were disrupted with QIAshredder homogenizer spin columns and an on-column DNase I digest was performed with the Qiagen RNase-Free DNase Set. RNA was eluted in 30 µl of 60 °C warm RNase-free molecular biology grade water. The RNA was then treated with 1 µl TURBO DNase (2 U/µl) for 1 h at 37 °C according to the manufacturer's instructions (Ambion via ThermoFisher Scientific, Germany). RNA samples were analysed for quantity and integrity on a Fragment Analyser capillary gel electrophoresis system (Advanced Analytical, USA) with a DNF-471 Standard Sensitivity RNA Analysis Kit. Six microlitres of each RNA sample were then reverse transcribed into cDNA using the Qiagen QuantiTect Reverse Transcription Kit according to the manufacturer's instructions. This protocol included an additional DNase treatment step and the reverse transcription reaction using a mix of random hexamers and oligo(dT) primers. Control reactions to test for gDNA contamination were done for all samples by adding ddH₂O instead of reverse transcriptase to the reaction mix. The cDNA was diluted twofold with RNase-free H₂O and analysed by qPCR with gene-specific primers. The qPCR reagents and conditions were the same as described above for genomic DNA qPCR. For data presentation, any qPCR reactions that yielded no C_q value after 40 PCR cycles were treated as C_q = 40. The no-template controls had an average C_q value of 39.16 with a standard deviation of 2.20.

Concentration, purification, and electron microscopy of reactivated mavirus particles. Five hundred millilitre cultures of strains E4-10P and E4-10M1 at 5×10^5 cells per millilitre in 3 l polycarbonate Fernbach flasks were either mock-infected with f/2 medium or infected with CroV at an MOI of 0.02. Six replicates were prepared for a total volume of 3 l per condition (E4-10P or E4-10M1, mock-infected or CroV-infected). At 3 d p.i., the cultures were centrifuged for 40 min at 7000g and 4 °C (F9 rotor, Sorvall Lynx centrifuge) and the supernatants were filtered on ice through a 0.2 µm PES Vivaflow 200 tangential flow filtration (TFF) unit (Sartorius via VWR, Germany). The filtrates were then concentrated on ice with a 100,000 MWCO PES Vivaflow 200 TFF unit to a final volume of approximately 15 ml. The concentrates were passed through a 0.1 µm pore-size PVDF Millex syringe filter (Millipore Merck, Ireland) and analysed on 1.1–1.5 g/ml continuous CsCl gradients. The CsCl gradients were prepared by underlaying 6.5 ml of 1.1 g/ml CsCl solution in 10 mM Tris-HCl, pH 8.0, 2 mM MgCl₂ with an equal volume of 1.5 g/ml CsCl solution in 10 mM Tris-HCl, pH 8.0, 2 mM MgCl₂ in a SW40 Ultra-Clear centrifuge tube (Beckman Coulter, Germany). Tubes were capped and continuous gradients were generated on a Gradient Master (BioComp Instruments, Canada)

with the following settings: tilt angle 81.5°, speed 35 r.p.m., duration 75 s. After replacing 3.9 ml of solution from the top of the gradients with 4 ml of concentrated culture supernatants, the gradients were centrifuged for 24 h, 205,000g, 18 °C using a SW40 rotor (Beckman Coulter, Germany) in a Beckman Optima ultracentrifuge. Bands in the gradients were visualized by illumination with an LED light source from the top of the gradient. One millilitre of gradient material from the mavirus band material (or equivalent positions of gradients were no such band was visible) was extracted with a syringe by puncturing the centrifuge tube with a 21-gauge needle. The extracted band material was dialysed for 24 h at 4 °C in 3 ml dialysis cassettes (Pierce, 20 kDa cutoff) against 1 l of 10 mM Tris-HCl, pH 8.0, 2 mM MgCl₂. After dialysis, each sample was diluted to 4 ml with 10 mM Tris-Cl, pH 8.0, 2 mM MgCl₂ and centrifuged in Ultra-Clear tubes (Beckman Coulter, Germany) in a SW60 rotor for 1 h, 100,000g, 18 °C. The supernatant was discarded and the pellets were softened overnight at 4 °C in 50 µl of 10 mM Tris-Cl, pH 8.0, 2 mM MgCl₂ and then re-suspended by pipetting. Aliquots (≈3 µl) of the concentrated samples were incubated for 2 min on Formvar/Carbon coated 75 mesh Cu grids (Plano, Germany) that had been hydrophilized by glow discharge. Grids were rinsed with ddH₂O, stained for 90 s with 1% uranyl acetate, and imaged on a Tecnai T20 electron microscope (FEI, USA) with an acceleration voltage of 200 kV.

Ultraviolet treatment of reactivated mavirus particles. A Stratalinker UV crosslinker 2400 (Stratagene) was used for irradiation of virus samples with ultraviolet-C (λ = 254 nm) light. Drops (500 ml) of 0.1-µm-filtered reactivated mavirus suspension were pipetted on Parafilm and irradiated with a single dose of 500 J m⁻² of ultraviolet-C light. The dose was monitored with a VLX-3W radiometer (Vilber-Lourmat). The irradiated virus suspension was then kept in the dark to prevent eventual light-induced DNA repair. Infection experiments were performed as described above and cultures were incubated in the dark for the entire duration of the experiment. Samples for DNA extraction and qPCR analysis were taken and processed as described above.

MiSeq and PacBio genome sequencing. Genomic DNA from 1×10^9 cells each of the clonal *C. roenbergensis* strains E4-10P and E4-10M1 was isolated using a Qiagen Blood & Cell Culture DNA Midi Kit. The genomes were sequenced on an Illumina MiSeq platform (Illumina, San Diego, California, USA) using the MiSeq reagent kit version 3 at 2 × 300-bp read length configuration. The E4-10P genome was sequenced by GATC Biotech AG (Constance, Germany) with the standard MiSeq protocol. The E4-10M1 genome was prepared and sequenced at the Max Planck Genome Centre (Cologne, Germany) with NEBNext High-Fidelity 2 × PCR Master Mix chemistry and a reduced number of enrichment PCR cycles (six) to reduce AT-bias. The total output was 6.8 Gbp and 4.5 Gbp for E4-10P and E4-10M1, respectively. Overall sequencing quality was assessed with FastQC version 0.11.3. Reads were trimmed for low quality bases and adapter contamination using Trimmomatic version 0.32 (ref. 22) and customized parameters (minimum phred score 20 in a 10-bp window, minimum length 75 bp, Illumina TruSeq3 reference adapter), resulting in 5.0 Gbp and 2.9 Gbp high-quality paired-end sequences, respectively. We also sequenced genomic DNA of strains E4-10P and E4-10M1 on a Pacific Biosciences RS II platform (two SMRT cells each, Max Planck Genome Centre, Cologne, Germany), which resulted in 0.52 Gbp and 1.30 Gbp of raw reads, respectively. The reads were extracted from the raw data files with DEXTRACTOR rev-844cc20 and general quality was assessed with FastQC version 0.11.3.

Read correction and assembly. Proovread version 2.12 (ref. 23) was used for hybrid correction of the PacBio reads with the respective trimmed MiSeq read sets using the following additional config settings: “seq-filter” = “> {‘--trim-win’ = > ‘10,1’, ‘--min-length’ = > 500}; sr-sampling’ = “> {DEF = > 0}”. Correction generated 423 Mbp (N50: 5994 bp) and 741 Mbp (N50: 7328 bp) of high-accuracy long reads for E4-10P and E4-10M1, respectively. Reads were assembled into contigs with SPAdes version 3.5.0 (ref. 24) using the dipsades.py module. Trimmed MiSeq reads were provided as paired-end libraries and corrected PacBio reads as single-end libraries. To account for structurally diverging sister chromosomes caused by asexual reproduction, the -expect-rearrangements flag was set in dipsades. Assembly metrics were assessed with QUAST version 2.3 (ref. 25). The E4-10P data set was assembled into 326 consensus contigs of at least 1,000 bp, with a total assembly length of 40.3 Mbp and an N50 of 290 kbp. The E4-10M1 genome was assembled into 463 consensus contigs longer than 1,000 bp, with a total assembly length of 31.4 Mbp and an N50 of 177 kbp.

Reference-guided assembly of the integrated mavirus genome. The E4-10M1 genome assembly was scanned for mavirus integration sites with blastn (NCBI BLAST version 2.2.29+ (ref. 26)). The search returned one partial hit with 7,000 bp and a few small hits with less than 600 bp alignment length. Additionally, partial hits were visualized and analysed in context of the assembly graph structure using Bandage version 0.4.2 (ref. 27). A full-length assembly of the potentially integrated mavirus genome sequence from the E4-10M1 set was generated through a reference guided assembly approach: corrected PacBio reads of the E4-10M1 strain were

aligned to the mavirus reference genome with blastn and strict settings (-evalue 10^{-10} -perc_identity 96). Matching reads longer than 1,000 bp were extracted and assembled with SPAdes version 3.5.0 (ref. 24) with the --only-assembler flag set.

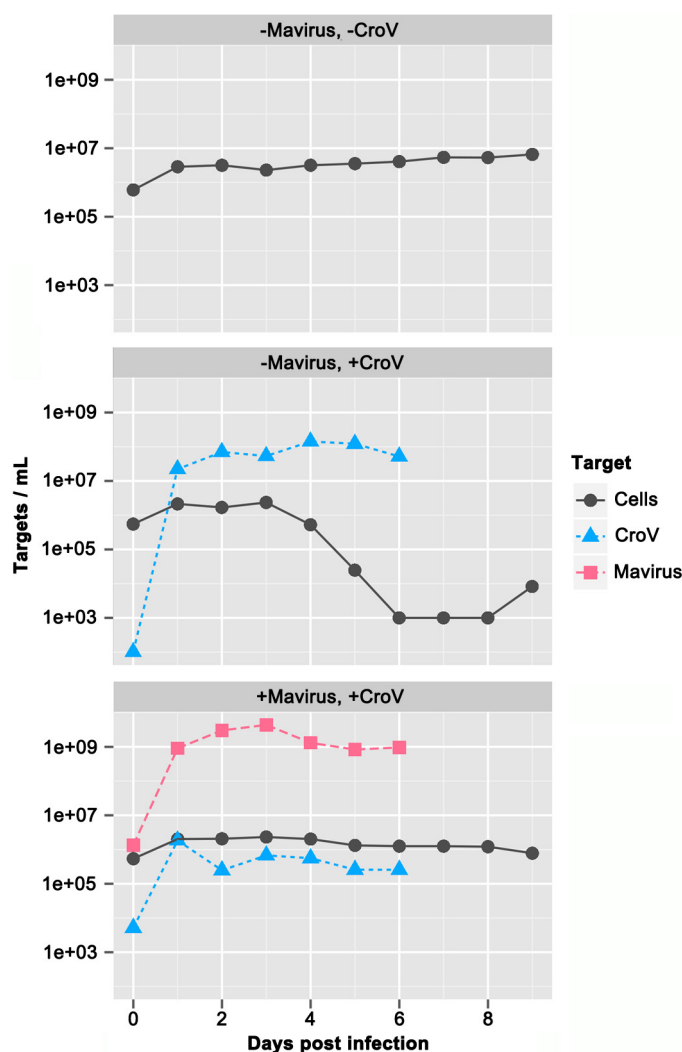
Detection/analysis of integration sites. Mavirus integration sites in the host genome were detected indirectly by identification of reads covering the junctions between a location in the *C. roenbergensis* genome and the terminal region of mavirus. In preparation, paired E4-10M1 MiSeq reads were merged with FLASH version 1.2.11 (ref. 28) into longer single-end fragments to maximize the chances for unambiguous hits in subsequent mappings. The merged fragments as well as the corrected E4-10M1 PacBio reads were aligned to the revised TIR region of the mavirus genome with bwa mem (BWA version 0.7.10-r984-dirty²⁹) and SAMtools version 1.1 (ref. 30). Fragments with a minimum alignment length of 30 bp and a minimum overlap of 10 bp at the TIR 5' end were identified and extracted with a custom script. Owing to the total length of 615/616 bp for the TIR, no merged MiSeq fragment spanned the entire TIR and, hence, no information about the strand-orientation of the mavirus core genome could be inferred from the MiSeq data. A read subset containing orientation information was generated by aligning extracted TIR-matching PacBio reads to the full mavirus genome and extracting end-overlapping reads with a minimum alignment length of 650 bp. These reads spanned the entire TIR and extended into one side of the core region by at least 34 bp, thus yielding information about the orientation of the integrated element. The extracted mavirus end-overlapping MiSeq and PacBio reads were mapped with bwa mem onto the E4-10M1 genome assembly. Mapping locations of the reads were considered potential integration sites and have been further analysed manually in a JBrowse³¹ genome browser instance, previously set up for the *C. roenbergensis* genome assemblies.

Reconstruction of a mavirus integration site. Direct assembly of an integrated mavirus genome into the host genome was prevented by the diploid state of the *C. roenbergensis* genome and by the repetitive nature of the multiple mavirus integrations, which could not be properly resolved in assembly graph structures. Therefore, we manually reconstructed a contig comprising a mavirus integration site from the previously obtained integration site coordinate information and read evidence available in the MiSeq and PacBio datasets. For the reconstruction, we chose the predicted integration site at nucleotide position 118,064 on contig 5 (length 208,205 bp). To validate the reconstructed sequence, MiSeq and corrected PacBio reads were mapped back against the artificial contig with bwa mem. Genomic features were annotated by mapping previously obtained host genome annotations (maker version 2.31.8)³² and mavirus gene annotations (PROKKA version 1.11 with custom mavirus database³³) onto the new contig. Annotations were mapped with a custom script based on University of California Santa Cruz (UCSC) annotation lift-over strategies (LiftOver_Howto, Minimal_Steps_For_LiftOver) using Kenttools version 302 (ref. 34). Visualization of the annotated contig was generated with bio2svg version 0.6.0.

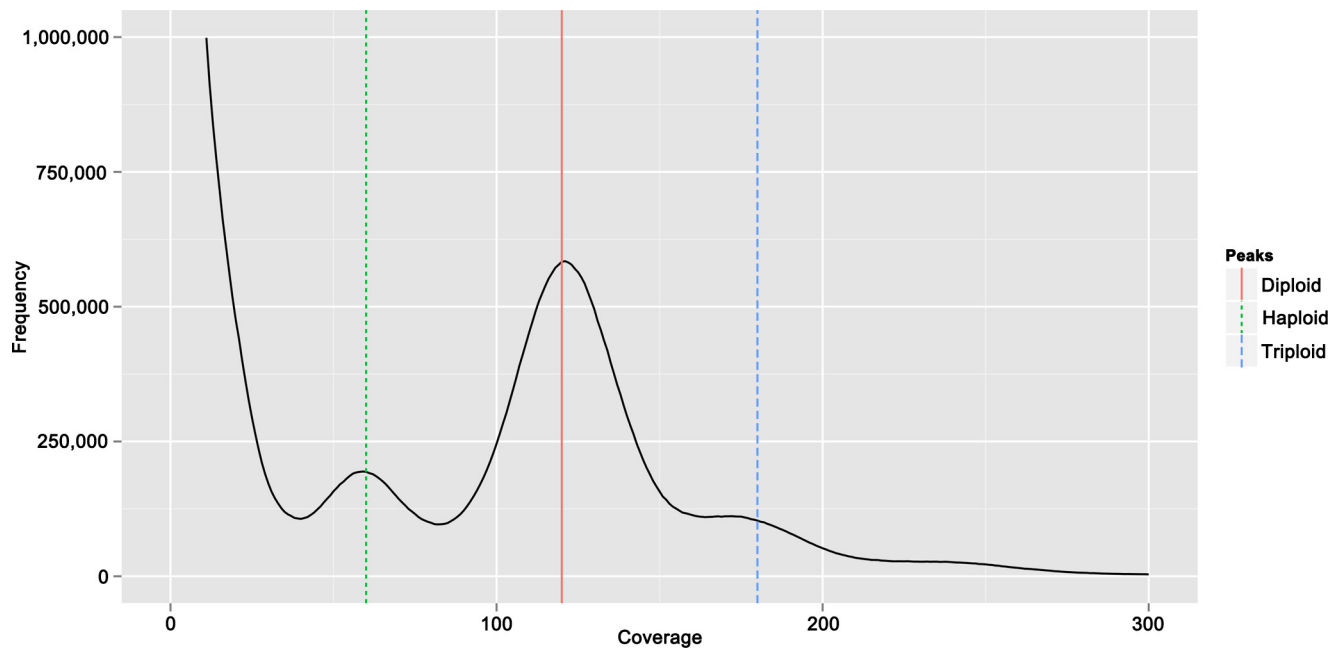
Ploidy assessment based on k-mer coverage frequency distribution. The 19-mer counts of the raw *C. roenbergensis* E4-10P Illumina MiSeq read dataset were calculated with jellyfish version 2.2.4 (ref. 35) in canonical representation and plotted with custom R scripts. Peak positions in Extended Data Fig. 2 were identified manually.

Data availability. *C. roenbergensis* strains E4-10P and E4-10M1 have been deposited in the Roscoff Culture Collection (strain numbers RCC 4624 and RCC 4625, respectively). The GenBank accession number for the reconstructed mavirus integration site of *C. roenbergensis* strain E4-10M1 shown in Fig. 1 is KU052222.

19. González, J. M. & Suttle, C. A. Grazing by marine nanoflagellates on viruses and virus-sized particles - ingestion and digestion. *Mar. Ecol. Prog. Ser.* **94**, 1–10 (1993).
20. Garza, D. R. & Suttle, C. A. Large double-stranded DNA viruses which cause the lysis of a marine heterotrophic nanoflagellate (*Bodo* sp) occur in natural marine viral communities. *Aquat. Microb. Ecol.* **9**, 203–210 (1995).
21. Reed, L. & Muench, H. A simple method of estimating fifty per cent end points. *Am. J. Hyg.* **27**, 493–497 (1938).
22. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
23. Hackl, T., Hedrich, R., Schultz, J. & Förster, F. proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* **30**, 3004–3011 (2014).
24. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
25. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
26. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
27. Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**, 3350–3352 (2015).
28. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
29. Li, H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30**, 2843–2851 (2014).
30. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
31. Skinner, M. E., Uzilov, A. V., Stein, L. D., Mungall, C. J. & Holmes, I. H. JBrowse: a next-generation genome browser. *Genome Res.* **19**, 1630–1638 (2009).
32. Campbell, M. S., Holt, C., Moore, B. & Yandell, M. in *Current Protocols in Bioinformatics* vol. 48, 4.11.1–4.11.39 (John Wiley, 2014).
33. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
34. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
35. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).

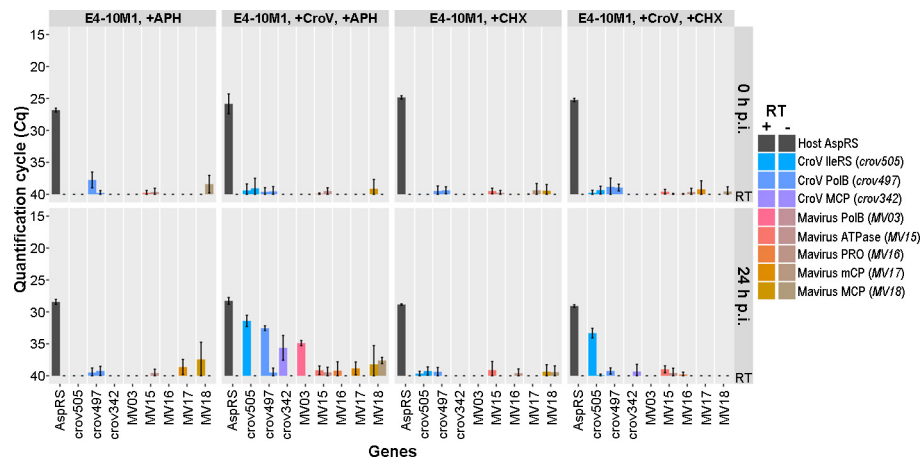


Extended Data Figure 1 | CroV-mavirus co-infection experiment of *C. roenbergensis* strain E4-10P to generate the mavirus-carrying strain E4-10M1. *C. roenbergensis* strain E4-10P was either mock-infected, infected with CroV, or co-infected with CroV and mavirus. Cell densities are based on microscopy counts and were monitored for 8 days, viral numbers were monitored for 6 days and are derived from qPCR data assaying short amplicons of the mavirus *MV18* gene (MCP) and the *croV283* gene (D11-like transcription factor), respectively. The detection limit for both methods was about 10^3 per millilitre. These experiments were performed in single copies. The mavirus-positive host strain E4-10M1 was isolated from the pool of surviving cells in the +mavirus, +CroV infection. See also Supplementary Spreadsheet 4.



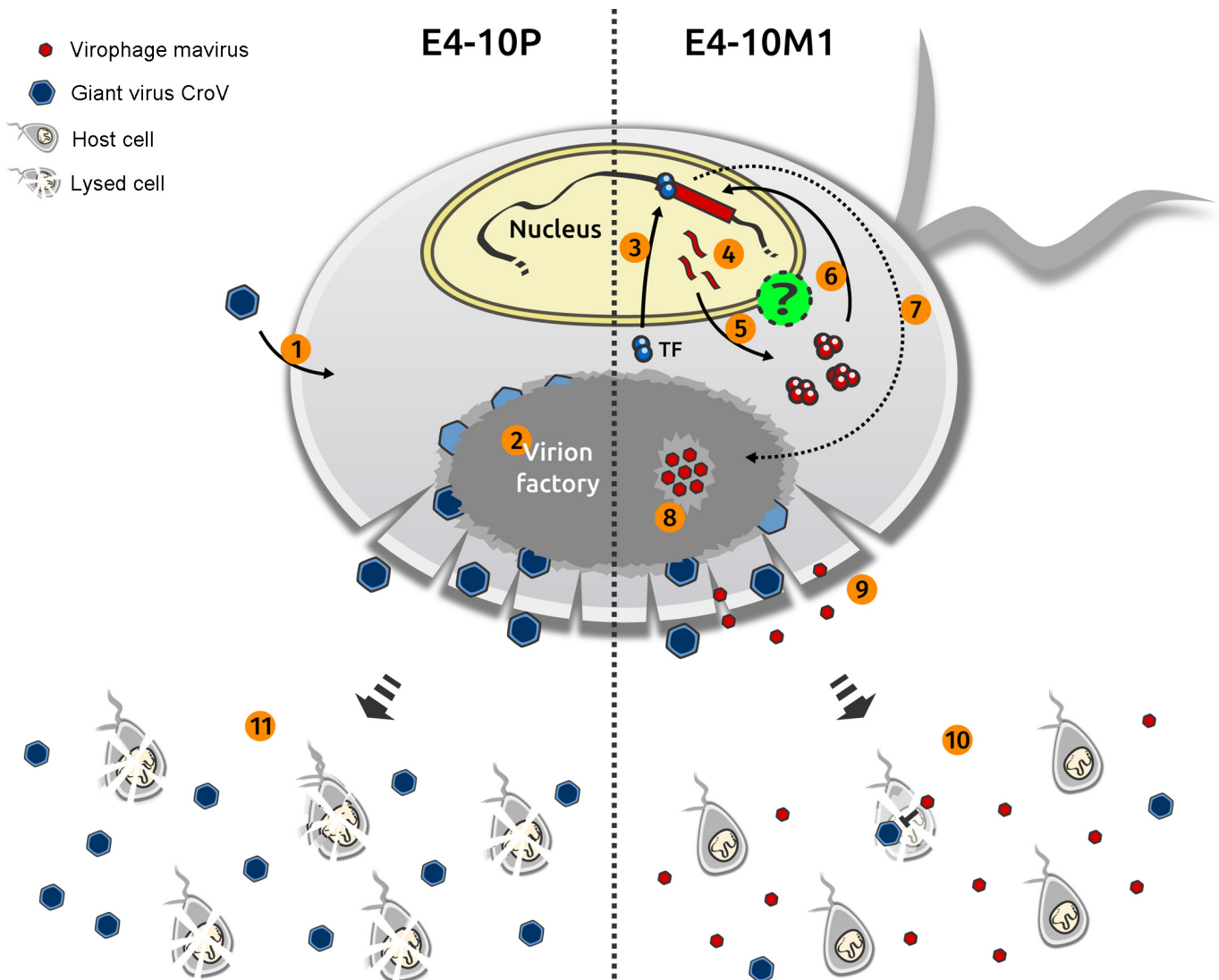
Extended Data Figure 2 | The *k*-mer frequency distribution of the diploid genome of *Cafeteria roenbergensis* strain E4-10P. Frequency distribution of *in silico* generated random 19-mers in the genomic read set of E4-10P. The distribution exhibits a major peak at 120× coverage (solid red line) corresponding to the majority of homozygous *k*-mers of the underlying diploid genome, a smaller peak at half the diploid coverage

(60×, dotted green line) comprising haplotype-specific *k*-mers, and a weak third peak at three times the haploid coverage (180×, dashed blue line) indicating a primarily diploid, partly triploid genome structure. Low-coverage *k*-mers derive from sequencing errors and bacterial contamination.



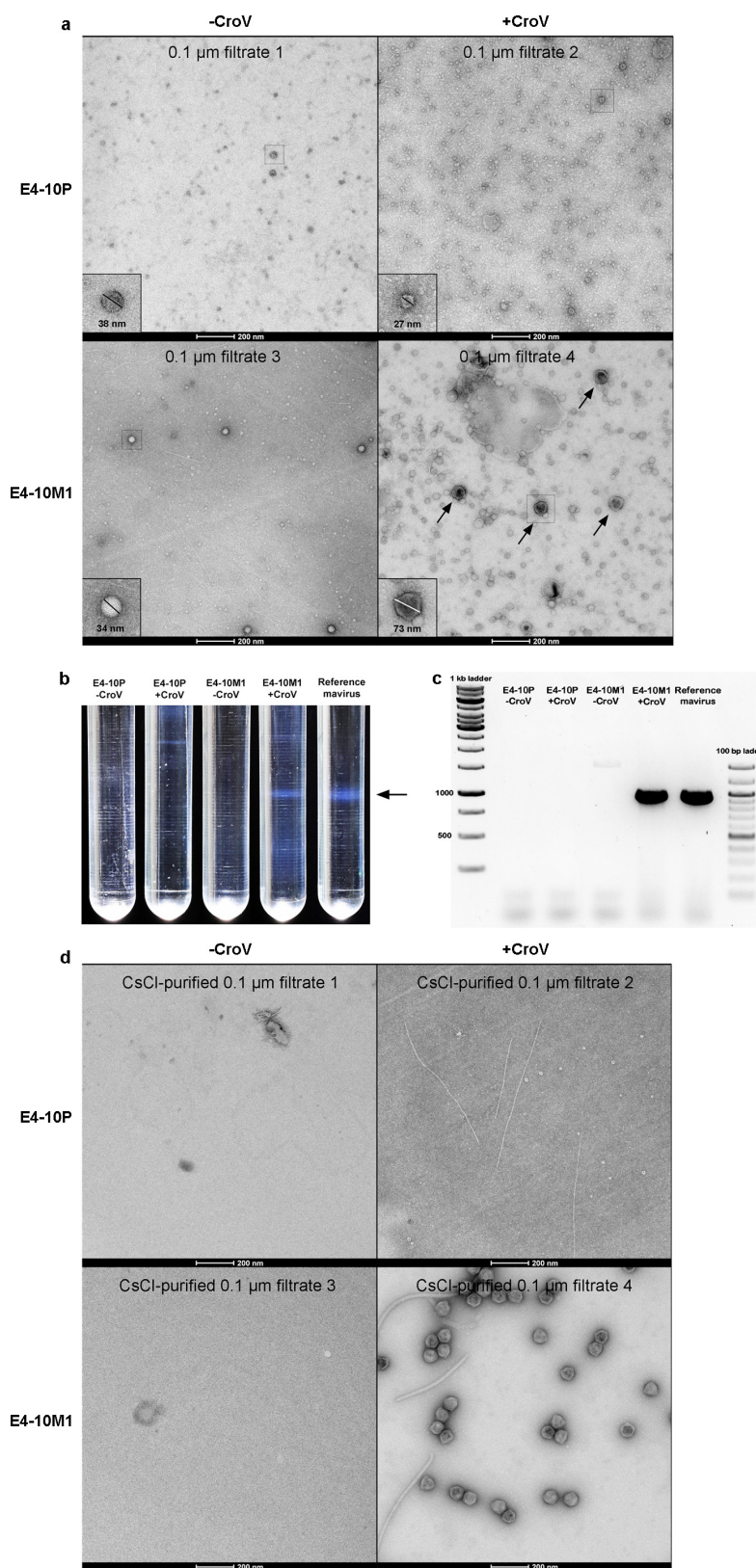
Extended Data Figure 3 | Gene expression of the endogenous mavirus genome is inhibited by cycloheximide (CHX) and aphidicolin (APH). Selected cellular and viral transcripts isolated at 0 and 24 h p.i. from mock-infected or CroV-infected E4-10P and E4-10M1 cultures in the presence of 5 μ g/ml aphidicolin or 50 μ g/ml CHX were quantified by qRT-PCR. Shown are the average quantification cycle (C_q) values of three independent experiments; error bars, s.d. The following genes were assayed: host *AspRS*, *C. roenbergensis* E4-10 aspartyl-tRNA synthetase; *crov342*, CroV

major capsid protein; *crov497*, CroV DNA polymerase B; *crov505*, CroV isoleucyl-tRNA synthetase; *MV03*, mavirus DNA polymerase B; *MV15*, mavirus genome-packaging ATPase; *MV16*, mavirus maturation protease; *MV17*, mavirus minor capsid protein; *MV18*, mavirus major capsid protein. C_q values of the reverse transcriptase-negative (–RT) reactions are shown directly to the right of the respective +RT results. Accession numbers are listed in Extended Data Table 3. See also Supplementary Spreadsheet 1.



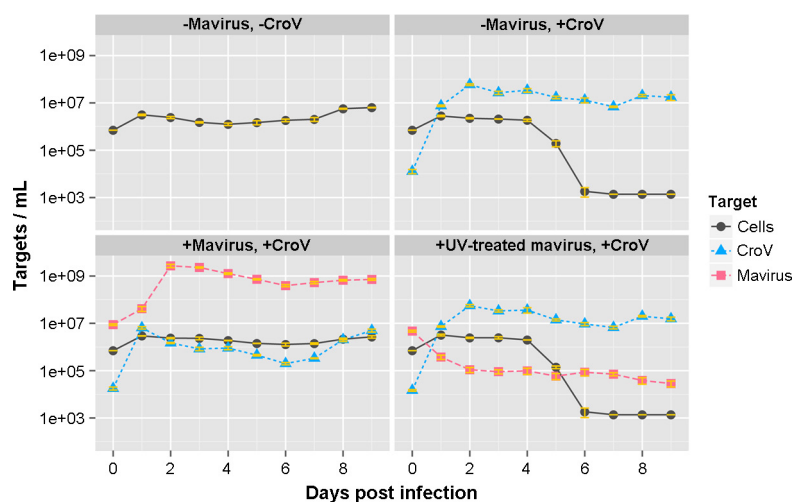
Extended Data Figure 4 | Hypothesis for CroV-induced reactivation of endogenous mavirus. Shown is a schematic *C. roenbergensis* cell displaying selected events of a CroV infection cycle in strains E4-10P (left) and E4-10M1 (right). Following CroV entry (1), the virion factory forms in the cytoplasm. At the onset of late phase, a CroV-encoded transcription factor recognizing the late CroV promoter motif is synthesized (2). We hypothesize that in E4-10M1 cells, the late transcription factor could enter the nucleus (3), bind the mavirus promoter sequences, and activate gene expression of the provirophage (4). Mavirus-specific transcripts would

then be exported and translated (5) and some of the mavirus proteins would return to the nucleus to excise or replicate the provirophage genome (6). The mavirus genome could then translocate to the CroV factory (7), where genome replication, particle assembly, and genome packaging would occur (8). Cell lysis releases the newly synthesized CroV and mavirus particles (9) and the reactivated virophages inhibit CroV propagation during subsequent co-infections (10). By contrast, CroV infection of an E4-10P cell does not induce a virophage response and CroV continues to infect other host populations (11).



Extended Data Figure 5 | Purification and characterization of reactivated mavirus particles. **a**, Three-litre cultures of CroV- or mock-infected E4-10P and E4-10M1 cultures were concentrated 200-fold and stained with uranyl acetate for electron microscopy. Representative particles from each filtrate are boxed and shown at higher magnification. Mavirus-like particles are marked by arrows. Filtrate numbers refer to the infections shown in Fig. 3a. **b**, The concentrated samples were analysed on 1.1–1.5 g/ml linear CsCl density gradients. A concentrated sample of reference mavirus was run in parallel and yielded a band at approximately

1.29 g/ml CsCl (arrow). Only the CroV-infected E4-10M1 culture produced a band at a similar density. **c**, PCR analysis of band material extracted from the CsCl gradients shown in **b** at a density of about 1.29 g/ml CsCl. Primers MaV21F and MaV21R were used to generate a 956-bp-long product of the *MV19* gene. **d**, Material from the 1.29 g/ml CsCl band or from equivalent positions was extracted from the gradients and visualized by negative-stain electron microscopy. Only the CroV-infected E4-10M1 culture contained mavirus-like particles.



Extended Data Figure 6 | Ultraviolet light treatment abolishes infectivity of reactivated mavirus. Reactivated mavirus contained in filtrate 4 from the infection experiments shown in Fig. 3a was irradiated with 500 J m⁻² of ultraviolet-C light ($\lambda = 254$ nm). Ultraviolet light-treated and untreated mavirus suspensions were tested for infectivity by

co-infection of host strain E4-10P with CroV. As shown in the lower right panel, reactivated mavirus treated with ultraviolet light was no longer able to replicate in the presence of CroV. Data are shown as the mean of biological triplicates \pm s.d. See also Supplementary Spreadsheet 5.

Extended Data Table 1 | Cell and virus concentrations in different size fractions of mock-infected and CroV-infected E4-10P and E4-10M1 populations

Condition	Host strain	Target	Unfiltered	5.0 µm filtrate (CN)	0.45 µm filtrate (PES)	0.22 µm filtrate (PES)	0.1 µm filtrate (PVDF)
Uninfected	E4-10P	Cells/mL	(1.53±0.18)E+06	(9.20±4.20)E+03	BDL	BDL	BDL
		Mavirus/mL	BDL	BDL	BDL	BDL	BDL
		CroV/mL	BDL	BDL	BDL	BDL	BDL
	E4-10M1	Cells/mL	(1.39±0.14)E+06	(5.50±0.00)E+03	BDL	BDL	BDL
		Mavirus/mL	(4.03±0.61)E+06	(1.04±0.46)E+04	BDL	BDL	BDL
		CroV/mL	BDL	BDL	BDL	BDL	BDL
CroV-infected	E4-10P	Cells/mL	BDL	BDL	BDL	BDL	BDL
		Mavirus/mL	BDL	BDL	BDL	BDL	BDL
		CroV/mL	(5.40±0.87)E+07	(1.35±0.10)E+06	(2.34±1.00)E+06	(3.11±0.47)E+04	(9.55±4.00)E+04
	E4-10M1	Cells/mL	BDL	BDL	BDL	BDL	BDL
		Mavirus/mL	(1.89±0.17)E+09	(1.25±0.05)E+09	(9.90±0.53)E+08	(8.97±0.71)E+08	(4.76±0.59)E+08
		CroV/mL	(1.41±0.12)E+08	(3.03±0.15)E+07	(1.00±0.03)E+07	(1.48±0.48)E+04	(1.18±0.20)E+04
Mavirus-spiked	E4-10P	Cells/mL	(1.83±0.19)E+06	(2.29±0.64)E+04	BDL	BDL	BDL
		Mavirus/mL	(1.22±0.02)E+09	(7.37±4.67)E+08	(7.64±4.52)E+08	(9.74±0.72)E+08	(7.23±0.14)E+08
		CroV/mL	BDL	BDL	BDL	BDL	BDL
Mechanically lysed	E4-10M1	Cells/mL	(1.83±1.41)E+04	BDL	BDL	BDL	BDL
		Mavirus/mL	(1.24±0.05)E+07	(1.15±0.04)E+07	(1.16±0.02)E+07	(1.17±0.05)E+07	(1.17±0.10)E+07
		CroV/mL	BDL	BDL	BDL	BDL	BDL

Uninfected host cultures or CroV-infected cultures after cell lysis were passed through syringe filters of various nominal pore sizes. As controls, E4-10P cells were spiked with mavirus particles immediately before filtration, and uninfected E4-10M1 cells were mechanically lysed by sonication and then filtered. DNA extracted from identical volumes of each filtrate was used as template in qPCR assays with mavirus- and CroV-specific primers. Cell concentrations were determined by microscopy counts. Shown are the average values of three independent experiments; error bars, s.d. *C. roenbergensis* cells are 5–10 µm in diameter, CroV particles are 300 nm in diameter, mavirus particles are 75 nm in diameter. BDL, below detection limit (~10³ cells or viruses per millilitre); CN, cellulose nitrate; PES, polyethersulfone; PVDF, polyvinylidene difluoride.

Extended Data Table 2 | Details on the 11 bioinformatically well-supported mavirus integration sites in *C. roenbergensis* strain E4-10M1

Site #	Contig length (bp)	Position of integration site within contig (bp)	TSD	# of terminal C/G nucleotides in the TIR
1	223481	121402	CGGAA	7
2	161079	145387	TGACAC	7
3	358887	10602	ATTTC	7
4	58090	3551	CAAAC	6
5*	208205	118064	GAGGCT	6
6	116230	100054	CGACA	7
7	14206	376	TGTCAA	6
8	209165	76864	CTGTG	7
9	403211	1650	CACCTC	7
10	19689	16490	CCACAC	7
11	8714	3914	TGCAC	7

The integration site described in detail in Fig. 1b is marked with an asterisk.

Extended Data Table 3 | PCR oligonucleotide primers used in this study

Primer name	5' to 3' sequence	Target sequence	Amplicon size	Application
Spezi-qPCR-1	ATGGTGCGACTTTTACTTATG	<i>MV03</i> pPolB	159 bp	qRT-PCR
Spezi-qPCR-2	AAGGGGTTATATTTTCCATTAG	(ADZ16402.1)		
MV15-qPCR-F	GTTATGAAATGGACGATGACG	<i>MV15</i> ATPase	123 bp	qRT-PCR
MV15-qPCR-R	CGCTAATAAAATATCTATCACTG	(ADZ16414.1)		
MV16-qPCR-F	TTTGATTCTTACGGAGGCAG	<i>MV16</i> PRO	172 bp	qRT-PCR
MV16-qPCR-R	TGGCACTGTTATTTGTTAGAG	(ADZ16415.1)		
MV17-qPCR-F	TACCTTTGTAGATGAAGATGG	<i>MV17</i> mCP	104 bp	qRT-PCR
MV17-qPCR-R	CTTCTTCTCCTATGATACCG	(ADZ16416.1)		
Spezi-qPCR-5	TAGTGGTGCTCTGGCTAATGGCTT	<i>MV18</i> MCP	125 bp	qPCR, qRT-PCR
Spezi-qPCR-6	TTAATCCCAGATCGGAAGGACGGA	(ADZ16417.1)		
CroV-qPCR-9	CTAAATTGGCCAGGTCTGGGTCTT	<i>crov283</i> D11-like TF	128 bp	qPCR
CroV-qPCR-10	CGTGGTAGAGTGGGTGAGAATGAA	(ADO67316.1)		
CroV-qPCR-11	CCTGAACTCACTAATGTATCC	<i>crov342</i> MCP	160 bp	qRT-PCR
CroV-qPCR-12	CATCTTTGAGAGGACGTAATTT	(ADO67376.1)		
CroV-qPCR-13	CTACAATTCTGCGGACTCTTC	<i>crov497</i> DNA PolB	166 bp	qRT-PCR
CroV-qPCR-14	CTTACAGGAGATGTGGCAGC	(ADO67531.1)		
CroV-qPCR-27	CATCAGAAGAACTTATTGGACA	<i>crov505</i> IleRS	153 bp	qRT-PCR
CroV-qPCR-28	CAAACACGATAATCATCTTCAC	(ADO67539.1)		
Cr_E4-10-AspRS-F	CATGGAGATGACCTTCAACG	<i>C. roenbergensis</i> AspRS	142 bp	qRT-PCR
Cr_E4-10-AspRS-R	GAAGTCCTCGTGGGGTACT			
CrE_cont6-3	TTGTGCGTGCTTCTCCAAG	<i>C. roenbergensis</i> strain	964 bp	PCR to verify integration site
CrE_cont6-6	GAAGGCGTTTCACTGTCCACT	E4-10P genome assembly		
MaV37	AGCACATACTCTCCAGAAAG	<i>MV01</i>	with CrE_cont6-3: 1979 bp	PCR to verify integration site
		(ADZ16400.1)		
MaV39	CACCACTCATTGAAGGTGAAGG	<i>MV19</i>	with CrE_cont6-6: 1502 bp	PCR to verify integration site
		(ADZ16418.1)		
MaV21F	GTTAGAGGCAACAGAGTTGG	<i>MV19</i>	956 bp	PCR to detect mavirus in
MaV21R	GCATTTTGTGCGGTTAATTCT	(ADZ16418.1)		CsCl gradient

GenBank accession numbers are listed where available.

Designed proteins induce the formation of nanocage-containing extracellular vesicles

Jörg Votteler¹, Cassandra Ogohara^{2,3}, Sue Yi^{2,3}, Yang Hsia^{2,3,4}, Una Nattermann^{2,3,4}, David M. Belnap^{1,5}, Neil P. King^{2,3} & Wesley I. Sundquist¹

Complex biological processes are often performed by self-organizing nanostructures comprising multiple classes of macromolecules, such as ribosomes (proteins and RNA) or enveloped viruses (proteins, nucleic acids and lipids). Approaches have been developed for designing self-assembling structures consisting of either nucleic acids^{1,2} or proteins^{3–5}, but strategies for engineering hybrid biological materials are only beginning to emerge^{6,7}. Here we describe the design of self-assembling protein nanocages that direct their own release from human cells inside small vesicles in a manner that resembles some viruses. We refer to these hybrid biomaterials as ‘enveloped protein nanocages’ (EPNs). Robust EPN biogenesis requires protein sequence elements that encode three distinct functions: membrane binding, self-assembly, and recruitment of the endosomal sorting complexes required for transport (ESCRT) machinery⁸. A variety of synthetic proteins with these functional elements induce EPN biogenesis, highlighting the modularity and generality of the design strategy. Biochemical analyses and cryo-electron microscopy reveal that one design, EPN-01, comprises small (~100 nm) vesicles containing multiple protein nanocages that closely match the structure of the designed 60-subunit self-assembling scaffold⁹. EPNs that incorporate the vesicular stomatitis viral glycoprotein can fuse with target cells and deliver their contents, thereby transferring cargoes from one cell to another. These results show how proteins can be programmed to direct the formation of hybrid biological materials that perform complex tasks, and establish EPNs as a class of designed, modular, genetically-encoded nanomaterials that can transfer molecules between cells.

Based on an advancing understanding of enveloped virus assembly, we proposed that synthetic proteins could be engineered to direct their own release from eukaryotic cells within membrane envelopes if they encoded three essential activities: membrane binding (termed ‘M’ domain activity), self-assembly (interaction or ‘I’ domain activity), and the ability to recruit ESCRT machinery to catalyse the final membrane fission step required for release from the cell (late budding or ‘L’ domain activity)^{8,10,11}. To test this hypothesis, we genetically fused peptide sequences capable of membrane binding and ESCRT recruitment to the computationally designed, 60-subunit nanocage I3-01 (ref. 9) to create the EPN-01 constructs (see Fig. 1a and Extended Data Fig. 1a for definitions of EPN-01 and EPN-01*). To promote membrane binding, we added to I3-01 an *N*-myristoylation signal corresponding to the first six amino acids of the HIV-1 structural Gag protein¹². To promote ESCRT recruitment, we added the 52-residue HIV-1 Gag p6 peptide (p6^{Gag}) to the C terminus^{8,11}. Size-exclusion chromatography and negative-stain electron microscopy of EPN-01* expressed in *Escherichia coli* cells (which lack *N*-myristoyltransferase and ESCRT factors and therefore produced non-enveloped nanocages) verified that purified EPN-01*, like the core I3-01 scaffold⁹, self-assembled into

regular nanocages approximately 25 nm in diameter (Extended Data Fig. 2).

When expressed in human embryonic kidney 293T cells, 13 ± 3% of the EPN-01 protein was released into the culture supernatant in complexes that could be pelleted by centrifugation through a 20% sucrose cushion (Fig. 1b). By contrast, a core I3-01–Myc scaffold lacking membrane-binding and ESCRT-recruiting elements was not released. Mutational analyses confirmed that each of the three design elements was required for EPN-01 release. Specifically, EPN-01 constructs with point mutations designed to disrupt myristoylation (ΔM)¹² or the designed assembly interface (ΔI)⁹ were not detectably released. Mutations designed to block ESCRT factor recruitment also inhibited EPN-01 release; mutation of the p6^{Gag} TSG101/ESCRT-I binding site (ΔL1) reduced EPN-01 release 18-fold, and mutation of the ALIX binding site (ΔL2) abrogated release entirely, either alone or in combination with the TSG101/ESCRT-I binding site mutation (ΔL1 + ΔL2). Similar results were obtained with EPN-01* analogues (Extended Data Fig. 1). These experiments demonstrate that all three

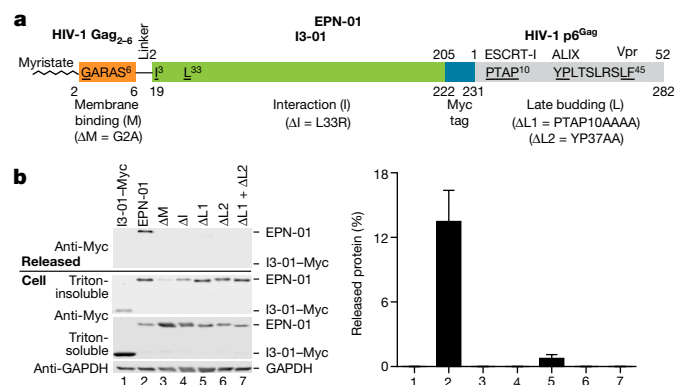


Figure 1 | EPN release requires three functional elements. **a**, Schematic of the EPN-01 construct, with the membrane-binding *N*-myristoylation element from HIV-1 Gag in orange (underlined G denotes the myristoylated glycine residue), the I3-01 self-assembly domain in green (the L33R mutation has previously been shown to prevent assembly⁹), the Myc epitope in blue, and the ESCRT-recruiting HIV-1 p6^{Gag} element in grey. The TSG101/ESCRT-I, ALIX- and Vpr-binding sites are shown and key interaction residues mutated in this study are underlined. Residue numbers above the construct correspond to individual functional elements, numbers below correspond to the overall construct. **b**, Left, western blots showing EPN-01 proteins harvested from 293T cell culture supernatants (top blot), cellular EPN-01 proteins in the Triton-insoluble and Triton-soluble fractions (middle blots), and a soluble cellular GAPDH loading control (bottom). To the right, the percentage of EPN-01 and mutants released into the supernatant is plotted (error bars show standard deviations from three technical repetitions).

¹Department of Biochemistry, University of Utah, Salt Lake City, Utah 84112, USA. ²Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. ³Institute for Protein Design, University of Washington, Seattle, Washington 98195, USA. ⁴Graduate Program in Biological Physics, Structure and Design, University of Washington, Seattle, Washington 98195, USA. ⁵Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA.

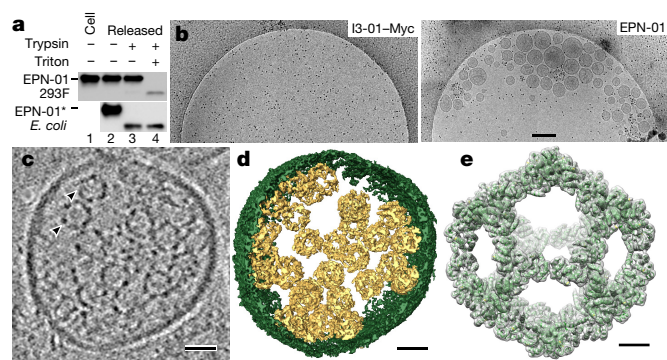


Figure 2 | EPNs comprise cell-derived membrane envelopes containing multiple protein nanocages. **a**, Top, western blot showing EPN-01 in 293F cells (lane 1) or released from 293F cells and treated as indicated (lanes 2–4). Bottom, western blot showing equivalent samples of non-enveloped EPN-01* nanocages purified from *E. coli*. **b**, Representative cryo-EM images showing extracellular vesicles/EPNs in culture supernatants from 293T cells that expressed I3-01-Myc (left) or EPN-01 (right). **c**, Central slice from a cryo-EM tomographic reconstruction of a released EPN; two internal protein nanocages are marked with arrowheads. **d**, Isosurface model of the 3D cryo-EM reconstruction from **c**. The EPN membrane is green and individual protein nanocages are gold. **e**, Single-particle cryo-EM reconstruction of the nanocages released from EPNs following detergent treatment. Charge density from the 5.7 Å resolution electron microscopy reconstruction is shown in grey (contoured at 4.5σ). The I3-01 computational design model⁹ (green ribbon) was fitted into the density as a rigid body. Scale bars, 300 nm (**b**), 25 nm (**c**, **d**) and 5 nm (**e**).

design elements—membrane binding, self-assembly and ESCRT recruitment—are functionally required for release of EPN-01 proteins from human cells.

Three different biochemical assays confirmed that released EPN-01 proteins were encapsulated within intact membrane envelopes. Figure 2a shows that released EPN-01 protein was protected against trypsin digestion in the absence of detergent, but became susceptible in the presence of 1% Triton X-100. In control experiments, non-enveloped EPN-01* nanocages produced in *E. coli* were degraded by trypsin in both the presence and absence of detergent. Detergent treatment was also required to render released EPN-01 proteins accessible to antibodies or to the small, polar aldolase substrate 2-keto-3-deoxy-6-phosphogluconate (KDPG), as assayed by I3-01 aldolase activity (Extended Data Fig. 3a, b). Thus, EPN-01 proteins were released within intact membrane envelopes that were impermeable to proteins or small polar molecules.

Several observations indicated that EPN-01 expression induced formation of extracellular vesicles. First, cryo-electron microscopy (cryo-EM) imaging revealed that culture supernatants from 293T cells expressing EPN-01* contained numerous vesicles, whereas cells that expressed the control I3-01-Myc did not (Fig. 2b). The vesicles averaged 107 ± 44 nm in diameter, and ranged between 40–320 nm (Extended Data Fig. 3c). Second, ALIX and actin, two cytosolic proteins that are often incorporated into enveloped viruses and exosomes^{13,14}, were released into the culture supernatants of cells that expressed EPN-01* (Extended Data Fig. 3d).

Cryo-EM tomography and 3D reconstructions revealed that each vesicle contained multiple protein nanocages that matched the known shape and size of I3-01 (ref. 9). The representative vesicle shown in Fig. 2c, d and Supplementary Video 1 is approximately 160 nm in diameter and contains 28 identifiable protein nanocages, most of which are associated with membranes, including a small internal vesicle. Reconstructions of EPN-01 vesicles of varying sizes revealed a correlation between vesicle size and nanocage number (Extended Data Fig. 3e).

To confirm that the internal protein nanocages corresponded to the design, they were released from the surrounding bilayer by detergent treatment, imaged by cryo-EM, and reconstructed as single

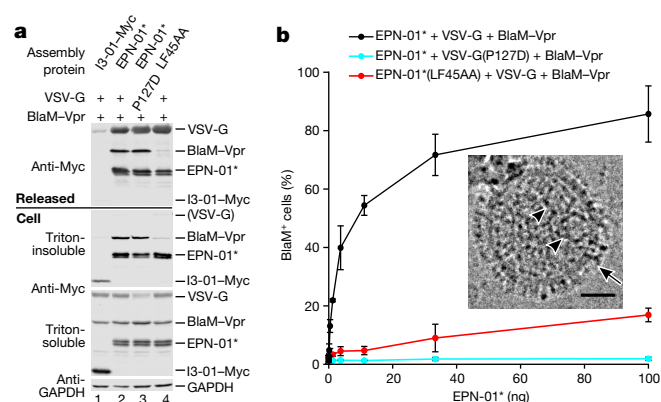


Figure 3 | EPNs can package macromolecular cargoes and deliver them into new target cells. **a**, Western blots showing cellular expression and release of I3-01-Myc, EPN-01*, and Myc-tagged VSV-G and BlaM-Vpr constructs (see Supplementary Table 3 for sequence information). Panels are equivalent to those in Fig. 1b. **b**, Cytoplasmic delivery to target HeLa cells by the three released constructs from **a**, plotted as percentages of BlaM-positive cells following incubation with the designated quantities of the different EPNs (error bars show standard deviations from two technical repetitions). Inset, cryo-EM image of an EPN-01*/VSV-G EPN with internal protein nanocages (arrowheads) and apparent VSV-G envelope spikes (arrow). Scale bar, 25 nm.

particles (Extended Data Fig. 4). A 5.7 Å resolution reconstruction of the released nanocages closely matched the computational I3-01 design model, including the computationally designed twofold interface between trimeric building blocks (Fig. 2e and Supplementary Video 2). Thus, EPN-01 protein expression induced the formation and release of vesicles that each contained multiple protein nanocages (termed EPNs).

Confocal immunofluorescence microscopy and immunogold electron microscopy of thin-sectioned cells expressing EPN-01 showed that the protein localized predominantly to internal compartments, but also to the plasma membrane (Extended Data Figs 5a, 6a, d). Control analyses of EPN-01(Δ M) confirmed that membrane localization was dependent upon myristoylation (Extended Data Figs 5b, 6b, d). Extracellular EPN-01 vesicles were observed and were decorated with anti-Myc immunogold particles as expected. We also occasionally observed immunogold-labelled vesicles that appeared to be in the process of budding from the plasma membrane (Extended Data Fig. 6c). Thus, vesicle biogenesis can apparently occur at the plasma membrane.

We next tested whether EPNs that incorporated the vesicular stomatitis viral glycoprotein (VSV-G) membrane fusion protein could deliver macromolecular cargoes into new target cells. To produce such EPNs, we co-expressed EPN-01* with VSV-G and with a protein comprising the β -lactamase (BlaM) enzyme fused to HIV-1 Vpr (BlaM-Vpr)^{15,16}. BlaM-Vpr was designed to be packaged within the EPN via the interaction of the Vpr domain with the ESCRT-recruiting p6^{Gag} polypeptide^{15,17} (Extended Data Fig. 7a).

EPN-01*, BlaM-Vpr and VSV-G were efficiently released together from 293T producer cells as designed (Fig. 3a). By contrast, cells that expressed the control I3-01-Myc construct released very little VSV-G and no BlaM-Vpr, indicating that these two proteins were probably released via EPN incorporation. Consistent with this interpretation, the fraction of EPN-01* assemblies with visible surface spikes increased fourfold upon VSV-G pseudotyping (Fig. 3b inset and Extended Data Fig. 7b). Two additional mutant constructs were used as controls to test whether the EPNs functioned as intended. In one case, the co-expressed VSV-G contained a mutation known to inhibit membrane fusion (VSV-G(P127D), Fig. 3a)¹⁸. In the other case, the Vpr binding site on the p6^{Gag} polypeptide in EPN-01* was mutated to inhibit Vpr binding (EPN-01*(LF45AA))¹⁷. As expected, both control constructs were released efficiently and the level of BlaM-Vpr packaging was

substantially reduced for EPN-01*(LF45AA). A similar dependence on the Vpr-p6^{Gag} interaction was observed for GFP-Vpr packaging (Extended Data Fig. 7c).

Released EPNs were harvested and tested for the ability to deliver their Vpr-BlaM cargoes into target HeLa cells. Delivery efficiencies were quantified by assaying cytosolic BlaM activity in target cells¹⁵ incubated with increasing concentrations of functional and mutant EPNs (Fig. 3b and Extended Data Fig. 7d). As designed, the EPNs with wild-type VSV-G and p6^{Gag} fused efficiently with target cells, with >80% of the target cells receiving Vpr-BlaM at the highest EPN-01* levels tested (Fig. 3b, black curve). By contrast, equivalent levels of either mutant EPN produced much lower percentages of BlaM-positive target cells (red and cyan curves). The slight activity of the EPN-01*(LF45AA) EPNs presumably reflected low residual levels of BlaM-Vpr packaging. The inactivity of the VSV-G(P127D)-containing EPNs confirmed that successful delivery required VSV-G-mediated membrane fusion to escape endocytic vesicles and access the cytoplasm.

To examine the modularity and generality of EPN design, we designed and tested a series of EPN constructs that carried a variety of membrane-binding, self-assembly and ESCRT-recruiting elements. Of the 43 EPN designs, 16 were robustly released from 293F cells within membrane-enclosed vesicles, as judged by the two criteria of detectable EPN release into the supernatant and protease susceptibility and enzymatic activity in the presence, but not absence, of detergent (Extended Data Fig. 8, Supplementary Tables 1 and 2). Seven additional designs showed weak but reproducible detergent-dependent enzymatic activity, while the remaining 20 designed EPN constructs failed to fulfil these criteria, indicating that they had design problems that prevented EPN formation. Nevertheless, the 16 successful designs indicate that the EPN design principles are robust and general.

The survey revealed that a variety of different membrane-binding, self-assembly and ESCRT-recruiting elements can function in EPN biogenesis. For example, a series of different membrane-binding domains from cellular proteins could substitute for the HIV-1 Gag myristoylation signal (Fig. 4a), including another fatty acid modification (palmitate) and a peripheral membrane-binding domain (the pleckstrin homology (PH) domain from phospholipase C δ). These membrane-binding domains differed in plasma membrane targeting efficiency (Extended Data Fig. 5), but in every case point mutations designed to disrupt membrane binding also blocked EPN release, demonstrating a functional requirement for membrane binding.

A second designed self-assembling scaffold that forms a 24-subunit assembly with octahedral symmetry (O3-33)³ could also support EPN formation when fused to an N-terminal p6^{Gag} peptide and a C-terminal PH domain. This construct was released efficiently, exhibited detergent-dependent protease sensitivity (Extended Data Fig. 8c), and again required membrane-binding, self-assembly and ESCRT-recruiting activities, as judged by reduced release of inactivating point mutants (Fig. 4b). Similarly, multiple different ESCRT-recruiting elements could also function in EPN release (Fig. 4c and Extended Data Fig. 8d). Examples included the equine infectious anaemia virus (EIAV) p9^{Gag} element, which recruits ALIX¹⁹, and an N-terminal element from Ebola virus VP40 protein, which recruits both TSG101/ESCRT-I and NEDD4 protein family members²⁰. Release was ESCRT-dependent in all cases because EPN production was strongly inhibited by overexpression of a dominant inhibitory version of the VPS4 ATPase that powers the ESCRT pathway²¹.

A remarkable property of enveloped viral structural proteins is that the membrane-binding, self-assembly and ESCRT-recruiting elements can often function from different positions within the polypeptide²², and this was also true of EPNs. For example, constructs in which the PH domain membrane-binding, the I3-01 self-assembly, and the HIV-1 p6^{Gag} ESCRT-recruiting elements were located at different positions in the protein sequence were released efficiently and in an ESCRT-dependent fashion (Fig. 4d and Extended Data Fig. 8e). Together with the EPN-18 construct described earlier (Fig. 4a), these experiments

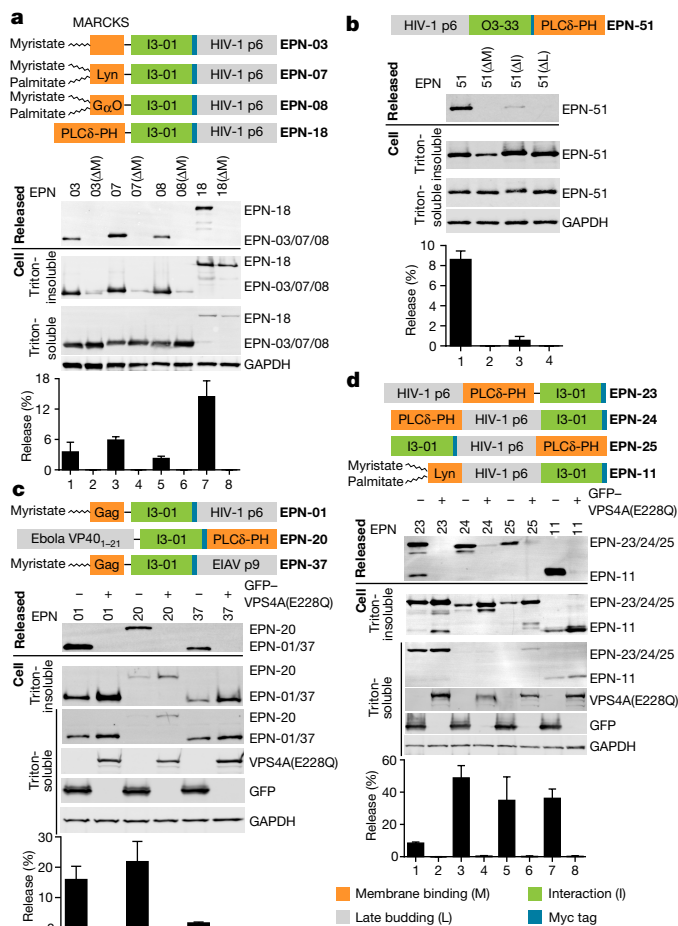


Figure 4 | A variety of functional elements and protein architectures support EPN formation. Schematic illustrations of constructs are shown, together with western blots and release quantification. **a**, Different membrane-binding domains support EPN release (lanes 1, 3, 5, 7). EPN point mutants shown in lanes 2, 4, 6 and 8 were designed to disrupt the membrane-binding interactions. **b**, The 24-subunit assembly O3-33 can function as an EPN self-assembly domain (lane 1). Mutants shown in lanes 2–4 were designed to disrupt the membrane-binding, self-assembly and ESCRT-recruiting elements. **c**, Different ESCRT-recruiting elements can support EPN release (lanes 1, 3, 5). ESCRT-dependent release is demonstrated by loss of EPN release upon co-expression of GFP-VPS4A(E228Q) (lanes 2, 4, 6). **d**, Membrane-binding, self-assembly and ESCRT-recruiting elements can function from different positions within EPN constructs (lanes 1, 3, 5, 7). ESCRT-dependent release is demonstrated by loss of EPN formation upon co-expression of GFP-VPS4A(E228Q) (lanes 2, 4, 6, 8). Error bars show standard deviations from three technical repetitions.

demonstrate that all three required elements can function from the middle, N or C termini of designed EPN proteins. Similarly, EPN-11 is a permutation of EPN-07 that is efficiently released and ESCRT-dependent (Fig. 4d and Extended Data Fig. 8e).

It was notable that EPN-01 induced the release of vesicles that typically contained multiple protein nanocages, whereas enveloped viruses usually contain a single capsid within their lipid bilayers (albeit with some interesting exceptions, for examples see ref. 23). However, the traditional demarcations between enveloped viruses, non-enveloped viruses and extracellular vesicles have become increasingly blurred, and there is now strong precedent for EPN-like assemblies in natural systems. One particularly well-characterized system is hepatitis A virus (HAV)²⁴. HAV has traditionally been categorized as a non-enveloped virus, but was recently shown to be released from cultured cells within small vesicles that typically contain multiple viral capsids²⁵. All of the

circulating HAV in patient serum is encapsidated within such vesicles, and their release is ESCRT-dependent. Thus, EPN biogenesis may mimic the natural process of HAV assembly and release. We note that simulations have suggested that membrane-curvature-driven effects promote nanocage clustering in vesicles (and vice versa)²⁶, and these effects probably contribute to EPN biogenesis. Going forward, we anticipate that the adaptable synthetic EPN systems described here will be useful in elucidating parameters that govern such processes as particle envelopment, membrane curvature, vesicle size, capsid numbers and target cell entry.

The central importance of biological membranes has inspired several recent approaches for designing biomimetic materials with membrane envelopes^{27,28} and engineering naturally occurring extracellular vesicles for therapeutic applications^{29,30}. Here we designed self-organizing protein nanocages that induce their own release inside cell-derived membrane envelopes. A key feature of our strategy is that it enables control over the biogenesis and contents of the materials through modification of EPN protein sequences. EPNs are highly modular and tolerant to substantial alterations, suggesting that they could be engineered to incorporate a wide variety of properties and functions tailored to desired applications.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 July; accepted 31 October 2016.

Published online 30 November; corrected online 7 December 2016

(see full-text HTML version for details).

- Rothmund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
- Ke, Y., Ong, L. L., Shih, W. M. & Yin, P. Three-dimensional structures self-assembled from DNA bricks. *Science* **338**, 1177–1183 (2012).
- King, N. P. et al. Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171–1174 (2012).
- Suzuki, Y. et al. Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* **533**, 369–373 (2016).
- Lai, Y. T. et al. Structure of a designed protein cage that self-assembles into a highly porous cube. *Nat. Chem.* **6**, 1065–1071 (2014).
- Mou, Y., Yu, J. Y., Wannier, T. M., Guo, C. L. & Mayo, S. L. Computational design of co-assembling protein-DNA nanowires. *Nature* **525**, 230–233 (2015).
- Delebecque, C. J., Lindner, A. B., Silver, P. A. & Aldaye, F. A. Organization of intracellular reactions with rationally designed RNA assemblies. *Science* **333**, 470–474 (2011).
- Votteler, J. & Sundquist, W. I. Virus budding and the ESCRT pathway. *Cell Host Microbe* **14**, 232–241 (2013).
- Hsia, Y. et al. Design of a hyperstable 60-subunit protein icosahedron. *Nature* **535**, 136–139 (2016).
- Wills, J. W. & Craven, R. C. Form, function, and use of retroviral gag proteins. *AIDS* **5**, 639–654 (1991).
- Weissenhorn, W., Poudevigne, E., Effantin, G. & Bassereau, P. How to get out: ssRNA enveloped viruses and membrane fission. *Curr. Opin. Virol.* **3**, 159–167 (2013).
- Göttlinger, H. G., Sodroski, J. G. & Haseltine, W. A. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc. Natl Acad. Sci. USA* **86**, 5781–5785 (1989).
- Choi, D. S., Kim, D. K., Kim, Y. K. & Ghoo, Y. S. Proteomics, transcriptomics and lipidomics of exosomes and ectosomes. *Proteomics* **13**, 1554–1571 (2013).
- Chertova, E. et al. Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. *J. Virol.* **80**, 9039–9052 (2006).
- Cavrois, M., De Noronha, C. & Greene, W. C. A sensitive and specific enzyme-based assay detecting HIV-1 virion fusion in primary T lymphocytes. *Nat. Biotechnol.* **20**, 1151–1154 (2002).
- Tobiume, M., Lineberger, J. E., Lundquist, C. A., Miller, M. D. & Aiken, C. Nef does not affect the efficiency of human immunodeficiency virus type 1 fusion with target cells. *J. Virol.* **77**, 10645–10650 (2003).
- Kondo, E. & Göttlinger, H. G. A conserved LXXLF sequence is the major determinant in p6^{gag} required for the incorporation of human immunodeficiency virus type 1 Vpr. *J. Virol.* **70**, 159–164 (1996).
- Fredericksen, B. L. & Whitt, M. A. Vesicular stomatitis virus glycoprotein mutations that affect membrane fusion activity and abolish virus infectivity. *J. Virol.* **69**, 1435–1443 (1995).
- Strack, B., Calistri, A., Craig, S., Popova, E. & Göttlinger, H. G. AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. *Cell* **114**, 689–699 (2003).
- Martin-Serrano, J., Zang, T. & Bieniasz, P. D. HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. *Nat. Med.* **7**, 1313–1319 (2001).
- Monroe, N. & Hill, C. P. Meiotic clade AAA ATPases: protein polymer disassembly machines. *J. Mol. Biol.* **428** (9 Pt B), 1897–1911 (2016).
- Parent, L. J. et al. Positionally independent and exchangeable late budding functions of the Rous sarcoma virus and human immunodeficiency virus Gag proteins. *J. Virol.* **69**, 5455–5460 (1995).
- Frank, G. A. et al. Maturation of the HIV-1 core by a non-diffusional phase transition. *Nat. Commun.* **6**, 5854 (2015).
- Wang, X. et al. Hepatitis A virus and the origins of picornaviruses. *Nature* **517**, 85–88 (2015).
- Feng, Z. et al. A pathogenic picornavirus acquires an envelope by hijacking cellular membranes. *Nature* **496**, 367–371 (2013).
- Reynwar, B. J. et al. Aggregation and vesiculation of membrane proteins by curvature-mediated interactions. *Nature* **447**, 461–464 (2007).
- Perrault, S. D. & Shih, W. M. Virus-inspired membrane encapsulation of DNA nanostructures to achieve *in vivo* stability. *ACS Nano* **8**, 5132–5140 (2014).
- Hu, C. M. et al. Nanoparticle biointerfacing by platelet membrane cloaking. *Nature* **526**, 118–121 (2015).
- EL Andaloussi, S., Mäger, I., Breakefield, X. O. & Wood, M. J. A. Extracellular vesicles: biology and emerging therapeutic opportunities. *Nat. Rev. Drug Discov.* **12**, 347–357 (2013).
- György, B., Hung, M. E., Breakefield, X. O. & Leonard, J. N. Therapeutic applications of extracellular vesicles: clinical promise and open questions. *Annu. Rev. Pharmacol. Toxicol.* **55**, 439–464 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported in part by Deutsche Forschungsgemeinschaft (DFG) Fellowship VO 1836/1-1 (to J.V.), an NIH Molecular Biology Training Grant (T32GM008268) (Y.H.), a PHS National Research Service Award (T32GM007270) from NIGMS (U.N.), grants from the Bill & Melinda Gates Foundation (OPP1118840) and Defense Advanced Research Projects Agency (W911NF-14-1-0162 and W911NF-15-1-0645) (N.P.K.), and NIH grants RO1 AI 51174 and P50 082545 (W.I.S.). We thank P. Shen and J. McCullough for assistance and advice on cryo-EM experiments, M. Kay and D. Eckert and J. Marvin (University of Utah FACS Core) for help with BlaM assays, S. Carter and S. Magesrawan (Caltech) for assistance with Amira Software, L. Nikolova (University of Utah EM Core) for assistance with immunogold labelling experiments, M. Redd (University of Utah Cell Imaging Core) for assistance with confocal microscopy, A. Wargacki for cloning assistance, S. Hauschka for tissue culture assistance, and M. Lajoie, L. Stewart and D. Baker for helpful discussions.

Author Contributions J.V., N.P.K. and W.I.S. designed and coordinated the study. J.V. performed EPN and cellular protein release assays, IP assays, BlaM delivery assays, confocal fluorescence microscopy, and immunogold EM. J.V. and D.M.B. performed cryo-electron microscopy and tomography, including single-particle reconstruction. C.O. and S.Y. designed EPN constructs and produced, purified, and analysed EPN proteins expressed in *E. coli*. C.O. performed EPN release assays, protease protection assays, and aldolase assays. Y.H. performed aldolase assays and purified I3-01 from *E. coli*. U.N. performed negative stain EM on proteins purified from *E. coli*. J.V., C.O., D.M.B., N.P.K. and W.I.S. interpreted data. J.V., N.P.K. and W.I.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.P.K. (neilking@uw.edu) and W.I.S. (wes@biochem.utah.edu).

Reviewer Information Nature thanks O. Farokhzad, J. Hurley and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment except when counting VSV-G spikes (Extended Data Fig. 7b).

Plasmids and antibodies. I3-01-Myc and EPN-01* mammalian expression constructs were generated by PCR amplification of the coding sequences from a pET29b-EPN-01* expression vector and transferred into a CMV-based mammalian expression vector (pCMV, DNASU ID: EvNO00601609 (ref. 31)) using the *NotI* and *XhoI* restriction sites. Plasmids for mammalian cell expression of EPNs -01 through -51 were constructed and inserted into pCMV using the *KpnI* and *XhoI* restriction sites by Gibson Assembly³² using synthetic DNA (Gen9). The *E. coli* expression plasmid for EPN-01* was constructed by adding the N- and C-terminal functional elements to the I3-01 sequence⁹ by PCR and inserting it by Gibson Assembly into pET29b digested with *NdeI* and *XhoI*. Mutations were introduced by round-the-horn site-directed mutagenesis or PCR amplification followed by Gibson Assembly as indicated in Supplementary Table 3. All constructs were verified by sequencing. A comprehensive list of all plasmids and coding sequences sources is provided in Supplementary Table 3. All of the plasmids have been submitted to the Addgene repository (<https://www.addgene.org/>). A comprehensive list of all antibodies, sources and dilutions is provided in Supplementary Table 4.

Protein expression in *E. coli* and purification. Expression plasmids were transformed into BL21(DE3) *E. coli* cells, and cells were grown in LB medium supplemented with 50 mg l⁻¹ kanamycin (Sigma) at 37 °C to an OD₆₀₀ of 0.8. Protein expression was induced by addition of 0.5 mM isopropyl-thio-β-D-galactopyranoside (Sigma) and allowed to proceed for 3 h at 37 °C before cells were harvested by centrifugation.

For the EPN-01* protein shown in Extended Data Figs 2c and 3a, b cells from a 1 l expression culture were lysed by sonication in 20 ml of 50 mM Tris pH 8, 250 mM NaCl, 20 mM imidazole, 2.5 mM MgCl₂, 0.5 mM CaCl₂, 1 mM DTT, 1 mM phenylmethanesulfonyl fluoride (PMSF) supplemented with 20 mg DNase (Sigma) and 2 mg of RNase (Qiagen), and the lysates were clarified by centrifugation for 25 min at 51,000g, 4 °C. Ammonium sulphate was added to the clarified lysate to 60% saturation, incubated at room temperature for 15 min, and the precipitate pelleted by centrifugation for 15 min at 51,000g, 4 °C. The pelleted protein was resuspended in 20 ml of 25 mM Tris pH 8, 150 mM NaCl, 5 mM EDTA, 1 mM DTT and heated for 10 min at 75 °C. The solution was clarified by centrifugation for 15 min at 51,000g, 4 °C, filtered with a 0.45 μm filter (EMD Millipore), and concentrated using a Centricon concentrator (EMD Millipore). The protein was then purified using a Superose 6 10/300GL column in the same buffer, the fractions pertaining to the nanocage peak centred around 12 ml were pooled and concentrated, and the protein refractionated using the Superose 6 10/300GL column equilibrated in 25 mM Tris pH 8, 150 mM NaCl, 5 mM EDTA supplemented with 0.75% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS). Nanocage peak fractions were again pooled and concentrated, and protein concentration determined using the BCA assay (ThermoFisher).

The proteins shown in Extended Data Fig. 2a, b were purified using a combination of immobilized metal affinity chromatography (IMAC) and size-exclusion chromatography using a Superose 6 10/300 GL column. *E. coli* cells were lysed by sonication in 25 mM TRIS pH 8.0, 250 mM NaCl, 1 mM DTT, 20 mM imidazole supplemented with 1 mM phenylmethanesulfonyl fluoride, and the lysates were cleared by centrifugation for 25 min at 51,000g, 4 °C and filtered through 0.22 μm filters (Millipore). The proteins were purified from the filtered supernatants by IMAC via linear gradient elution from HisTrap HP columns (GE Healthcare) using 25 mM TRIS pH 8.0, 250 mM NaCl, 1 mM DTT, 20 mM imidazole as running/wash buffer and 25 mM TRIS pH 8.0, 250 mM NaCl, 1 mM DTT, 500 mM imidazole as elution buffer. Elution fractions containing pure proteins of interest were pooled, concentrated using centrifugal filters (Sartorius Stedim Biotech), and further purified on a Superose 6 10/300 gel filtration column (GE Healthcare) using 25 mM TRIS pH 8.0, 150 mM NaCl, 1 mM DTT as running buffer.

Negative stain electron microscopy. For the negative stain electron microscopy image shown in Extended Data Fig. 2c, 6 μl of purified EPN-01* at 0.075 mg ml⁻¹ were applied to glow discharged, carbon-coated 400-mesh copper grids (Ted Pella), washed with Milli-Q water and stained with 0.75% uranyl formate. Grids were visualized for assembly validation and optimized for data collection. Screening and sample optimization was performed on a 100 kV Morgagni M268 transmission electron microscope (FEI) equipped with an Orius charge-coupled device (CCD) camera (Gatan). The final image was recorded on a 120 kV Tecnai G2 Spirit transmission electron microscope (FEI) using an Ultrascan 4000 4k × 4k CCD camera (Gatan) at 52,000× magnification at the specimen level.

Mammalian cell culture. HeLa and HEK293T (293T) cells were obtained from ATCC and cultured in D-MEM (ThermoFisher) containing 10% FBS,

penicillin (100 U ml⁻¹) and streptomycin (0.1 mg ml⁻¹), at 37 °C and 5% CO₂. Expi293F (293F) cells, used to survey different EPN constructs, were obtained from ThermoFisher and cultured in 293F Expression Medium (ThermoFisher) containing penicillin (100 U ml⁻¹) and streptomycin (0.1 mg ml⁻¹), at 37 °C and 5% CO₂ while shaking at 125 r.p.m. Cells were tested for mycoplasma contamination every 3 months using the MycoAlert Mycoplasma Detection Kit (Lonza).

EPN release assays. To assay EPN release as shown in Figs 1, 4 and Extended Data Fig. 1, 8 × 10⁵ 293T cells were seeded in 6-well plates 24 h before transfection. Cells were transfected with 2 μg of plasmid DNA expressing the I3-01- or O3-33-based EPN constructs, or co-transfected with 1 μg of plasmids expressing I3-01-based constructs and 1 μg of either pEGFP-VPS4A(E228Q) or pEGFP-C1 (Clontech), using the polyethylenimine (PEI, Polysciences, 3 μl of PEI per μg DNA) method. The medium was replaced with 1 ml growth medium 5 h later. Cells and culture supernatants were harvested 24 h post transfection. Released EPN assemblies were collected from the culture supernatants by centrifugation through a 200 μl 20% sucrose cushion for 90 min at 21,000g, 4 °C, and denatured by adding 50 μl 1 × Laemmli buffer and boiling for 5 min. Cells were lysed for 5 min on ice in 200 μl cold lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1% Triton X-100, protease inhibitors). Lysates were clarified by centrifugation for 5 min, 16,000g, 4 °C. The Triton-soluble fraction was treated with 200 μl 2 × Laemmli buffer supplemented with 10% 2-mercaptoethanol (Sigma) and boiled for 5 min. Triton-insoluble material was solubilized in 200 μl 2 × Laemmli buffer by boiling for 10 min. Samples containing EPN-51 and mutants thereof were incubated for 30 min at 40 °C instead of boiling because the protein aggregated at high temperatures. Benzonase (Sigma, 0.5 μl per sample) was added to remove nucleic acids from the Triton-insoluble fractions. The Triton-soluble and -insoluble cellular fractions, and the released EPN complexes, were separated by 12.5% SDS-PAGE, transferred onto PVDF membranes (or nitrocellulose membranes in case of EPN-51 constructs), and probed with antibodies against the Myc epitope (primary antibodies and dilutions are provided in Supplementary Table 4). GAPDH was used as a loading control. Bands were visualized by probing the membrane with fluorescently labelled secondary antibodies (Li-Cor Biosciences) and scanning with an Odyssey Imager (Li-Cor Biosciences). Levels of expressed and released Myc-tagged EPN proteins were quantified by western blot densitometry with ImageJ³³ using standard curves generated with known quantities of recombinant EPN-01* protein produced in *E. coli*. Release efficiencies are reported as the percentage of EPN protein pelleted from the supernatant versus the total protein in cells and pelleted supernatant. All experiments were repeated independently at least twice. Standard deviations shown in all figures were calculated from three technical repeats (three transfections in parallel) of each experiment.

EPN purification. Released EPNs used in immunoprecipitation and cryo-EM studies were purified from culture supernatants of 293T cells (2 × 10⁶ per 10 cm plate, 7 plates per specimen for the experiments shown in Fig. 2b and Extended Data Fig. 3d, 36 plates for the experiments shown in Fig. 2c, e and Extended Data Fig. 3a, seeded 24 h before transfection) after transient transfection with plasmids encoding EPN-01 or EPN-01* (12 μg per plate) using the calcium phosphate method (Clontech). Transfected cells were incubated overnight and the media was replaced with exosome production media (D-MEM supplemented with 10% FBS, depleted of contaminating extracellular particles by centrifugation overnight at 100,000g at 4 °C and subsequently filtered through a 0.22 μm filter)³⁴. Cells were grown for an additional 24 h and extracellular EPN assemblies were purified by a series of filtering and centrifugation steps (adapted from ref. 34). In brief, cell debris was removed by centrifugation of the supernatant at 1,000g for 5 min followed by filtration through a 0.22 μm filter (EMD Millipore). EPN assemblies were collected by centrifugation at 100,000g in an SW32Ti (BeckmanCoulter) at 4 °C for 1 h. Pellets were resuspended in PBS and pooled in one tube (SW41 rotor, BeckmanCoulter). PBS was added to fill the tube completely and EPN assemblies were collected by centrifugation at 100,000g at 4 °C for 1 h. Pellets were resuspended in 1 ml of PBS and concentrated by centrifugation at 100,000g at 4 °C for 1 h in an OptimaMAX-E (BeckmanCoulter) bench-top ultracentrifuge using a TLS-55 rotor. EPNs were quantified by western blotting as described above. Typical yields were 2–8 μg EPN-01 and EPN-01* proteins from 36 × 10 cm dishes.

Preparation of EPNs -01 through -51 for protease protection and enzyme assays. EPNs for protease protection and aldolase activity assays as shown in Fig. 2a and Extended Data Figs 3b and 8 were prepared as follows. On the day of transfection, 293F cell count and viability were determined using trypan blue solution in a haemocytometer. The cells were plated in 1 ml volumes at 2.5 × 10⁶ cells per ml on non-TC treated 12-well plates (Corning). The cells were transfected with 1 μg of plasmid DNA using Expifectamine transfection reagent (ThermoFisher) following the manufacturer's instructions. A cocktail of Expifectamine 293F Transfection Enhancer1 and Enhancer2 (ThermoFisher) was added to each well following manufacturer's instructions 18 h after transfection. Cells and cultured supernatants

were collected 44 h post-transfection and separated by centrifugation for 5 min at 1,000g, 4 °C. The culture supernatants were filtered through a 0.45 µm filter (EMD Millipore) into a 1.5 ml microfuge tube. Released EPN assemblies were collected from the culture supernatants by centrifugation through 200 µl of a 20% sucrose cushion for 120 min at 21,000g, 4 °C and resuspended in PBS.

Protease protection assays. Purified EPNs were resuspended and incubated under three different conditions, 10 µl each: untreated EPN, EPN + 0.05 mg ml⁻¹ trypsin, and EPN + 0.05 mg ml⁻¹ trypsin + 1% Triton X-100. Samples were incubated for 30 min at 25 °C and then 1 mM PMSF was added and incubated for 10 min at 25 °C to inactivate trypsin. Samples were denatured by boiling for 10 min in 4 × Laemmli buffer supplemented with 5% 2-mercaptoethanol (except samples with the O3-33 domain, which were not boiled). All fractions were separated by SDS-PAGE, transferred onto nitrocellulose membranes, and analysed by western blot using an anti-Myc antibody (Supplementary Table 4). Western blots were imaged using HRP-conjugated secondary antibodies (Cell Signaling Technology) and Clarity Western ECL Blotting Substrate (Bio-Rad). At least three biological replicates of the protease protection assay (independent transfections or batches of *E. coli* purified EPN-01*) were performed for each construct described in Extended Data Fig. 8 and Supplementary Table 1.

Aldolase enzyme activity assays. The 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase activity of the I3-01 domain was monitored using a L-lactic acid dehydrogenase (LDH)-coupled assay³⁵. 95-µl samples of assay solutions containing 25 mM HEPES pH 7.0, 20 mM NaCl, 0.1 mM NADH, 0.11 U µl⁻¹ LDH, 1 mM KDPG, and either including or omitting 1% Triton X-100, were mixed with 5 µl of resuspended EPNs. Loss of absorbance at 339 nm owing to oxidation of NADH was monitored using a SpectraMax M3 plate reader. At least three biological replicates of the aldolase activity assay (independent transfections or batches of *E. coli* purified EPN-01*) were performed for each construct described in Extended Data Fig. 8 and Supplementary Table 1.

EPN immunoprecipitation assays. A total of 1 µg of purified EPN-01 expressed in either bacteria or harvested from 293T supernatants was incubated in 250 µl PBS buffer containing either 0.1% or 0.5% CHAPS detergent for 20 min. EPN-01 assemblies were immunoprecipitated by addition of 30 µl of the indicated antibodies coupled to agarose resin and incubated for 14 h at 4 °C on a rotating shaker. Antibody-bound resins were: anti-rabbit-IgG-Agarose (Sigma) and anti-c-Myc-Agarose (Sigma). Resins were washed six times at 4 °C with 1 ml PBS/0.1% CHAPS buffers and resuspended in 250 µl in 1 × Laemmli buffer containing 10% 2-mercaptoethanol, boiled for 5 min, and analysed by western blotting.

Cryo-EM tomographic imaging of EPNs. To prepare samples for cryo-EM tomography, 3 µl of purified EPNs in PBS (50 ng EPN-01 per µl as determined by western blotting versus a standard curve) were mixed with 3 µl of BSA-coated gold fiducials (10 nm size, Electron Microscopy Sciences). 3.5 µl of the suspension were applied to a glow-discharged R2/2 holey carbon coated EM grid (Quantifoil) within the environmental chamber of a Vitrobot (FEI) maintained at 4 °C, 80% relative humidity. Excess liquid was blotted for 7.5 s (0 mm offset) from the grids with filter paper (Whatman) before plunge freezing in liquid ethane. Cryo-grids were placed in a Gatan 626 cryoholder (Gatan) and imaged in a 200 kV Tecnai F20 microscope (FEI) equipped with a K2 Summit direct electron detector (Gatan). Tilt series were recorded bidirectionally starting from 0° to ± ~60° with a 1° step size at a magnification of 22,500× and a defocus of -8 µm (total dose per specimen, ~300 e⁻ Å⁻²) using low-dose mode in SerialEM³⁶. Tomograms were generated using the IMOD software package³⁷. Image stacks were aligned, binned by 4 and gold particles were erased using findbeads3d within IMOD. Aligned image stacks were Fourier filtered (cut-off 0.25, σ = 0.08) and tomographic reconstructions were performed using the simultaneous reconstruction technique (SIRT). Noise reduction was performed with the nonlinear anisotropic diffusion (NAD) method in IMOD³⁷, using a *K* value of 0.04 with 12 iterations. Segmentation and isosurface rendering was done using Amira (Version 4.1.2, FEI). Individual particles that could be completely traced along slices though the *z* axis were manually identified and surrounded by a mask. The space inside the mask was then segmented and an isosurface was generated. Video segments were created in Amira and ffmpeg (http://ffmpeg.org) was used to combine segments and generate the H.264 encoded Supplementary Video 1.

Cryo-EM imaging and reconstruction of EPN-01* nanocages. 4 µg of EPNs in PBS purified as described in EPN purification were incubated with 0.75% CHAPS for 20 min at 4 °C (20 µl total volume). 3.5 µl samples were placed on glow discharged R2/2 holey carbon grids (Quantifoil) within the Vitrobot environmental chamber (maintained at 4 °C, 80% relative humidity), blotted for 11 s (0 mm offset) with filter paper, and plunge frozen in liquid ethane. Cryo-grids were imaged with a TF20 microscope operated at 200 kV (42,000× magnification and -0.7 to -3.3 µm defocus). Images were recorded on a Gatan K2 Summit direct electron detector. SerialEM³⁶ was used to facilitate low-dose imaging and semi-automated

data collection, and 60 frames were recorded of each view in super-resolution mode. Frames were aligned and summed by using MotionCor2³⁸.

Three-dimensional images were reconstructed via routines implemented in the package SCIPION³⁹. Contrast transfer function (CTF) parameters were determined using CTFFIND4 (ref. 40). Particle images were selected and extracted using Xmipp^{41,42}. Non-dose-weighted image sums from MotionCor2 were used for CTF determination and particle picking, with dose-weighted image sums used for all other steps. Extracted two-dimensional particle images were processed and then classified with the RELION software package⁴³. Suitable two-dimensional class averages were used to determine an *ab initio* 3D model via the program RANSAC⁴⁴. This model was then used as the starting model for 3D image reconstruction via RELION with icosahedral symmetry applied during the 3D reconstruction calculations. The resulting 3D map was masked and a B-factor was applied (post-processing) via an automated procedure in RELION⁴³. The final map was constructed from 8,573 particles, and the resolution was determined to be 5.7 Å by the gold standard 0.143 criterion (Extended Data Fig. 4c). The design model⁹ was rigid-body fit into the completed 3D density, using UCSF Chimera⁴⁵ to perform the fit and generate Supplementary Video 2. One additional residue (Lys2 of I3-01) was built into the density at the N-terminal end of the first helix of the I3-01 construct, and no additional density was visible for any of the remaining sequences outside the I3-01 domain. The resolution of the model/map fit was 7.1 Å (Fourier cross resolution 0.378 criterion⁴⁶, see Extended Data Fig. 4c).

Confocal immunofluorescence microscopy. For immunofluorescence imaging shown in Extended Data Fig. 5, 2 × 10⁵ HeLa cells were seeded onto coverslips in 12-well plates and transfected with plasmids encoding EPN proteins the next day. 24 h post-transfection, cells were fixed with 3.5% paraformaldehyde in PBS for 10 min at room temperature, washed twice in PBS, blocked and permeabilized in a block/perm solution (0.1% Triton X-100, 3% BSA in PBS) for 10 min before primary anti-Myc antibodies were added (1 µg ml⁻¹ in block/perm solution) and incubated for 1 h at room temperature. Cells were washed three times for 10 min in wash buffer (0.1% Triton X-100 in PBS), and Alexa 488-labelled secondary anti-mouse IgG antibody was added (ThermoFisher, 2 µg ml⁻¹ in block/perm solution) and incubated for 1 h at room temperature. Nuclei and actin were stained by incubating cells with Hoechst 33342 (ThermoFisher, 1:10,000) and Phalloidin 647 (ThermoFisher, 1:40) in block/perm solution for 15 min. Cells were washed four times in wash buffer, and twice in PBS before mounting onto glass coverslips (Fluoromount-G, Southern Biotech). Confocal immunofluorescence images were acquired using NIS Elements software on a Nikon A1 microscope. Final images were prepared in Image J (FIJI)³³.

Immunogold labelling. For immunogold labelling experiments shown in Extended Data Fig. 6, 8 × 10⁵ 293T cells were seeded in 6-well plates 24 h before transfection (2 wells per sample). Cells were transfected with 2.5 µg of plasmid DNA expressing the EPN construct using Lipofectamine 2000 (ThermoFisher) following the manufacturer's instructions. 24 h post-transfection, the media was removed, cells were knocked loose and washed off the plate in fixative (2% PFA/0.1% glutaraldehyde in PBS), transferred to 1.5-ml test tubes and incubated on a rocker for 16 h at 4 °C. Cells were then pelleted for 4 min at 16,000g and washed four times for 5 min in 1 ml PBS and twice for 5 min in 1 ml water on a rocker. Cell pellets were stained with 50 µl 2% uranyl acetate for 30 min and washed again three times for 5 min in 1 ml water. Samples were dehydrated in 1 ml of a graded series of ethanol in water (3 × 70% ethanol, 3 × 95% ethanol, 3 × 100% ethanol, 5 min each) and incubated 16 h in 1 ml of a 1:1 mixture ethanol:LR White (Sigma-Aldrich) at room temperature. Samples were then infiltrated by two 6 h incubations in 1 ml 100% LR White, and the resin was polymerized at 50 °C overnight. Thin sections (80 nm, cut by a diamond knife (Diatome) in a Leica EM UC6 ultratome (Leica)) were mounted onto support specimen nickel grids (Electron Microscopy Sciences). For immunogold labelling, grids were hydrated on 100-µl drops of PBS for 10 min, and reactive aldehydes were then deactivated by incubating grids on 100-µl drops of 50 mM glycine in PBS for 10 min. Grids were blocked for 1 h on 40-µl drops of blocking solution (5% BSA in PBS) and then incubated with primary anti-Myc antibodies (40-µl drops, 2 µg ml⁻¹ in blocking solution) overnight at 4 °C in a moist chamber. The next day, grids were washed three times for 5 min on 100-µl drops of blocking solution and probed on 40-µl drops of secondary anti-mouse IgG antibody labelled with 10-nm gold particles (Ted Pella, 1:100 in blocking solution) for 2 h at room temperature. Grids were then washed three times for 5 min in blocking solution and three times for 5 min in water, dried, and viewed on a JEOL JEM1400 electron microscope at an accelerating voltage of 120 kV.

To quantify immunogold staining, a region of interest was defined in ImageJ and particles were counted inside the defined area using the particle analyse function (see Extended Data Fig. 6 a–c, right panels)³³.

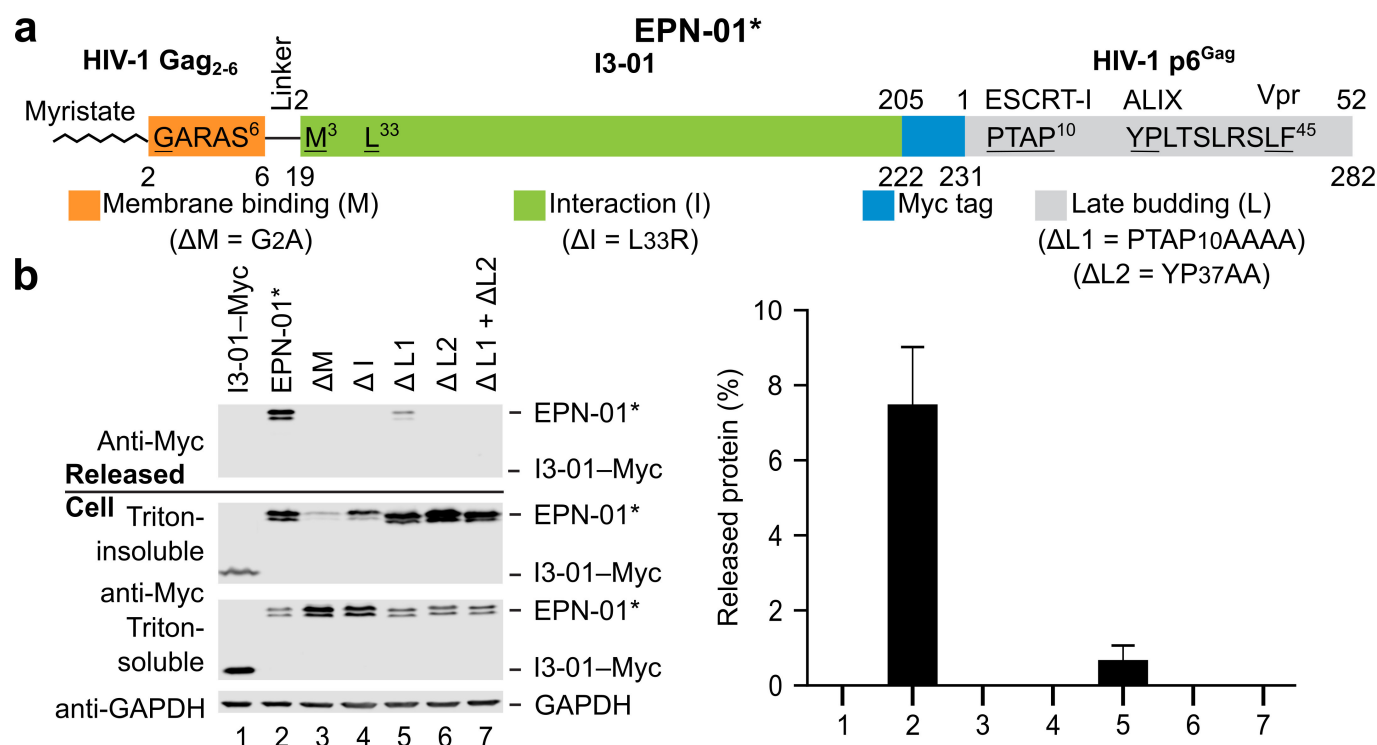
EPN-01 delivery assays. The ability of EPNs to deliver packaged enzymes into the cytoplasm of recipient cells, as shown in Fig. 3 and Extended Data Fig. 7, was evaluated using a modified version of the β -lactamase (BlaM) assay described previously¹⁵. 5 μ g of a plasmid expressing an N-terminally Myc-tagged chimaeric BlaM–Vpr fusion protein was co-transfected with 9 μ g EPN-encoding plasmids and 1 μ g of a plasmid encoding a C-terminally Myc-tagged VSV-G (Supplementary Table 3), in 10-cm plates using the Lipofectamine (ThermoFisher) method. Transfection medium was replaced with 10 ml growth medium 5 h post-transfection. EPNs containing VSV-G and BlaM–Vpr were collected 36 h post-transfection by centrifugation through 2 ml of a 20% sucrose cushion at 100,000g for 1 h at 4 °C in a SW-41 rotor (BeckmanCoulter). Cellular and released proteins were separated on a 12.5% SDS–PAGE and quantified by western blotting using an anti-Myc antibody and EPN-01 proteins purified from *E. coli*. The internal volume of each protein nanocage is around 3,000 nm³ (ref. 9), which is sufficient to package about 60 close-packed BlaM–Vpr molecules, assuming that each BlaM–Vpr molecule is approximately 50 nm³. Quantification of western blot band intensities indicated that an average of approximately 10 BlaM–Vpr molecules were actually packaged by each 60-subunit EPN-01* nanocage.

For the BlaM delivery assay, 2×10^5 HeLa cells per well were seeded in 24-well plates. 24 h later, the indicated EPN quantities were added to cultures and incubated for 2 h at 37 °C. EPN-containing supernatants were replaced by CCF2-AM labelling media, prepared according to the manufacturer's instructions (ThermoFisher) using CO₂-independent media (ThermoFisher) as the loading solution. Cells were labelled for 16 h at 13 °C and assayed by flow cytometry (FACSCanto, BD Biosciences) for changes in fluorescence emission spectrum from green (520 nm) to blue (447 nm). Data were collected with FACSDiva and analysed with FlowJo software (Treestar). Non-transduced cells treated with CCF2 were used to set the gate for uncleaved CCF2, which was set to discriminate transduced and non-transduced cells at a tolerance of <0.2% false positives (Extended Data Fig. 7d). Transduction assays were repeated independently at least three times (independent transfections). Standard deviations shown in Fig. 3 were calculated from three technical repeats of the BlaM delivery assay.

Data availability. Raw scans of all membranes and gels shown in the manuscript are included in this article as Supplementary Fig. 1. All other raw data are available from the corresponding authors upon request. Electron microscopy charge density maps, model fitting, and supporting data have been deposited in the EMDataBank under the accession number EMD-8278 (PDB accession

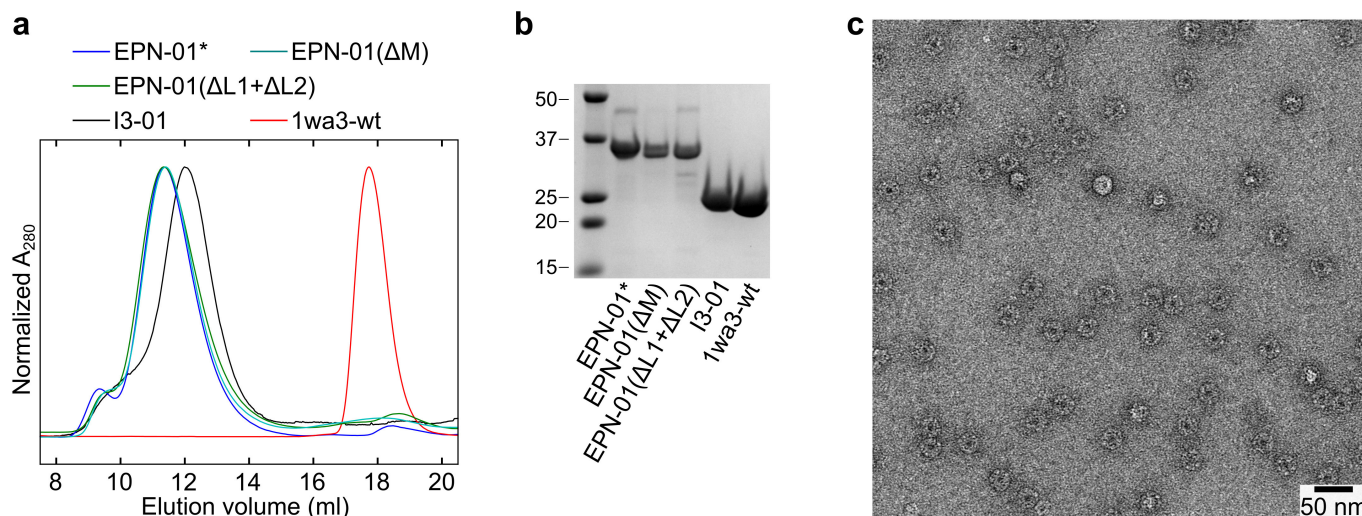
number 5KP9). All novel plasmid constructs (Supplementary Information, Supplementary Table 3) have been submitted and are accessible through the Addgene plasmid repository (<https://www.addgene.org/>).

31. Morita, E., Arai, J., Christensen, D., Votteler, J. & Sundquist, W. I. Attenuated protein expression vectors for use in siRNA rescue experiments. *Biotechniques* **0**, 1–5 (2012).
32. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
33. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
34. Thery, C., Amigorena, S., Raposo, G. & Clayton, A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Curr. Protoc. Cell Biol.* Chapter 3, Unit 3 22 (2006).
35. Griffiths, J. S. *et al.* Cloning, isolation and characterization of the *Thermotoga maritima* KDPG aldolase. *Bioorg. Med. Chem.* **10**, 545–550 (2002).
36. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
37. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
38. Zheng, S. Q., Palovcak, E., Armache, J., Cheng, Y. & Agard, D. A. Anisotropic correction of beam-induced motion for improved single-particle electron cryo-microscopy. Preprint at <http://dx.doi.org/10.1101/061960> (2016).
39. de la Rosa-Trevin *et al.* Scipion: a software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99 (2016).
40. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
41. Abrishami, V. *et al.* A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs. *Bioinformatics* **29**, 2460–2468 (2013).
42. de la Rosa-Trevin, J. M. *et al.* Xmipp 3.0: an improved software suite for image processing in electron microscopy. *J. Struct. Biol.* **184**, 321–328 (2013).
43. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
44. Vargas, J., Álvarez-Cabrera, A. L., Marabini, R., Carazo, J. M. & Sorzano, C. O. Efficient initial volume determination from electron microscopy images of single particles. *Bioinformatics* **30**, 2891–2898 (2014).
45. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
46. Penczek, P. A. Resolution measures in molecular electron microscopy. *Methods Enzymol.* **482**, 73–100 (2010).



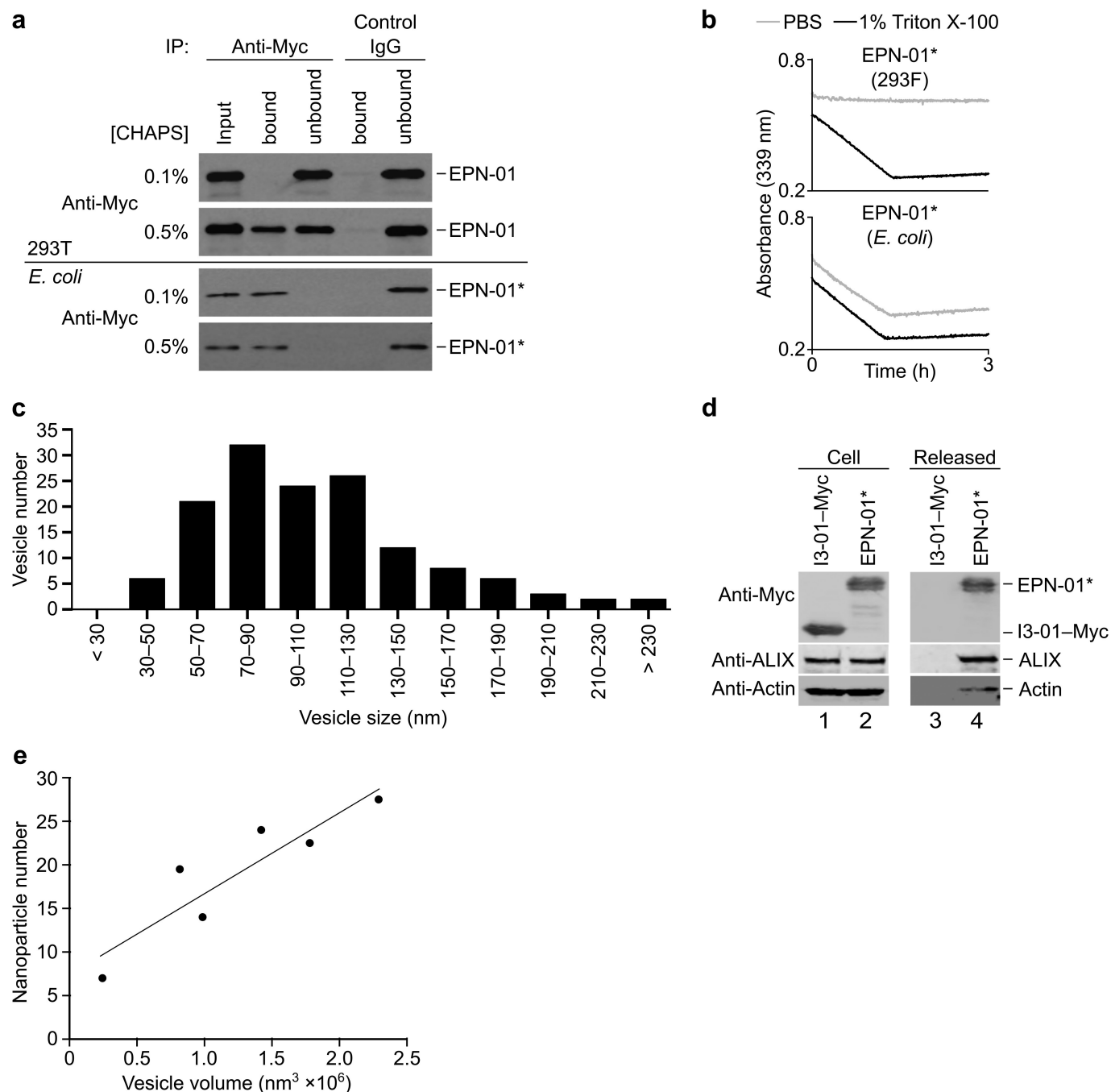
Extended Data Figure 1 | EPN-01* is released from cells. **a**, Schematic of the EPN-01* constructs. Note that the EPN-01* constructs shown in this figure are analogous to the EPN-01 constructs shown in Fig. 1 except that EPN-01* ran as a doublet because a methionine codon at I3-01 position 3 (isoleucine in EPN-01) served as a second translation initiation codon. **b**, Western blots showing functional and mutant EPN-01* proteins harvested from 293T cell culture supernatants (top panel). The middle panels are western blots showing cellular EPN-01* proteins in the Triton-

insoluble fraction (stably bound to membranes) and the Triton-soluble fraction (cytoplasm and weakly bound to membranes), respectively. The bottom panel is a western blot showing a GAPDH loading control from the Triton-soluble fraction. Right, the percentage of each protein released into the supernatant is plotted (error bars show standard deviation from three technical repetitions). In summary, EPN-01 and EPN-01* were functionally equivalent and were employed interchangeably in our studies. For western blot source data, see Supplementary Fig. 1.



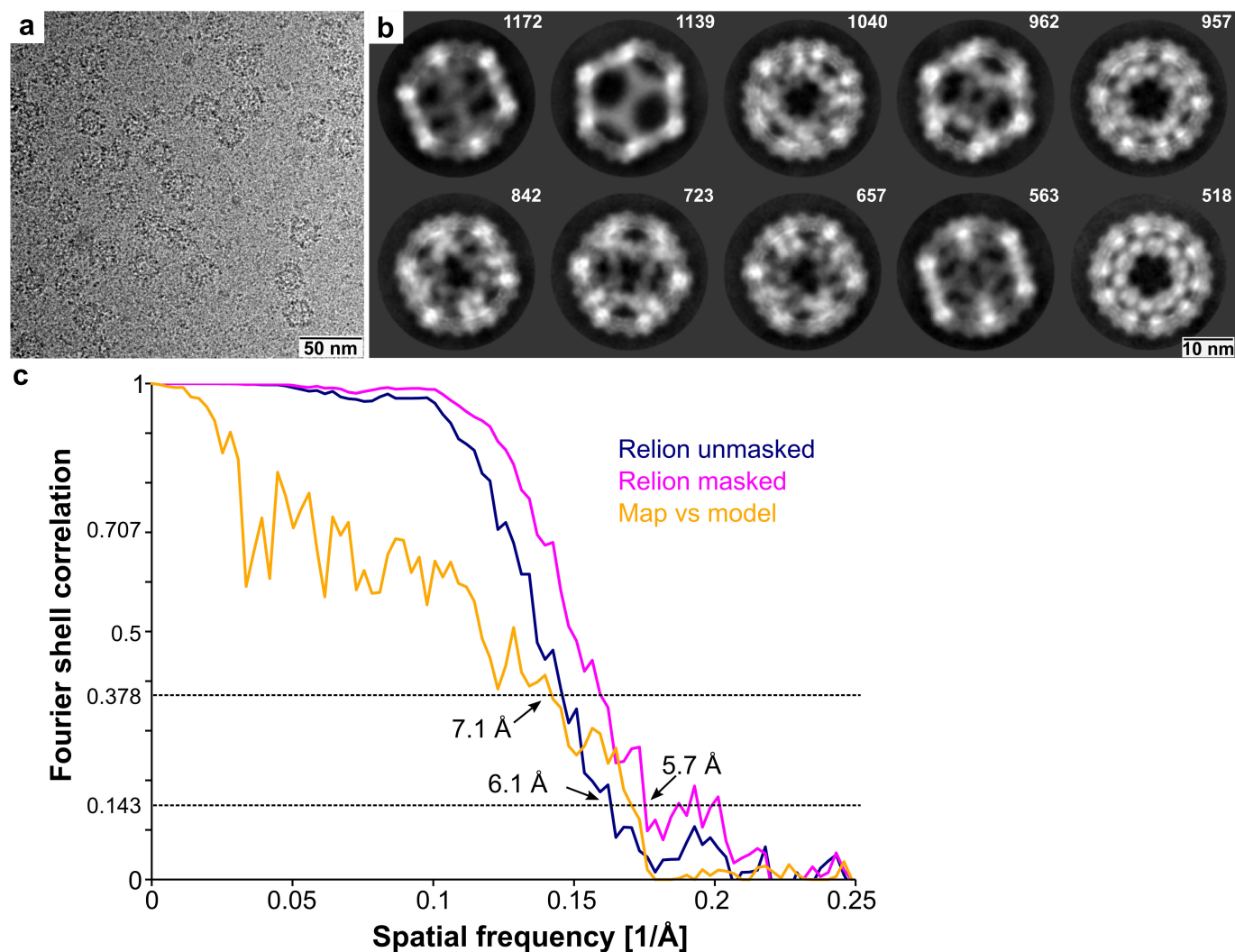
Extended Data Figure 2 | EPN-01* forms nanocages that resemble those of the I3-01 scaffold. **a**, Size-exclusion chromatograms (Superose 6 10/300 GL) of EPN-01*, EPN-01(Δ M), EPN-01(Δ L1 + Δ L2), I3-01, and 1wa3-wt purified from *E. coli*. 1wa3-wt is the wild-type trimeric aldolase from which I3-01 was derived⁹. All of the proteins eluted as nanocages

except for 1wa3-wt, which eluted at the volume expected for the trimer. **b**, SDS-PAGE of the peak fraction from each of the chromatograms in **a**. Molecular weight marker positions are indicated in kilodaltons. **c**, Negative stain electron microscopy image of purified *E. coli* EPN-01*, showing a homogeneous field of particles (\sim 25 nm diameter).



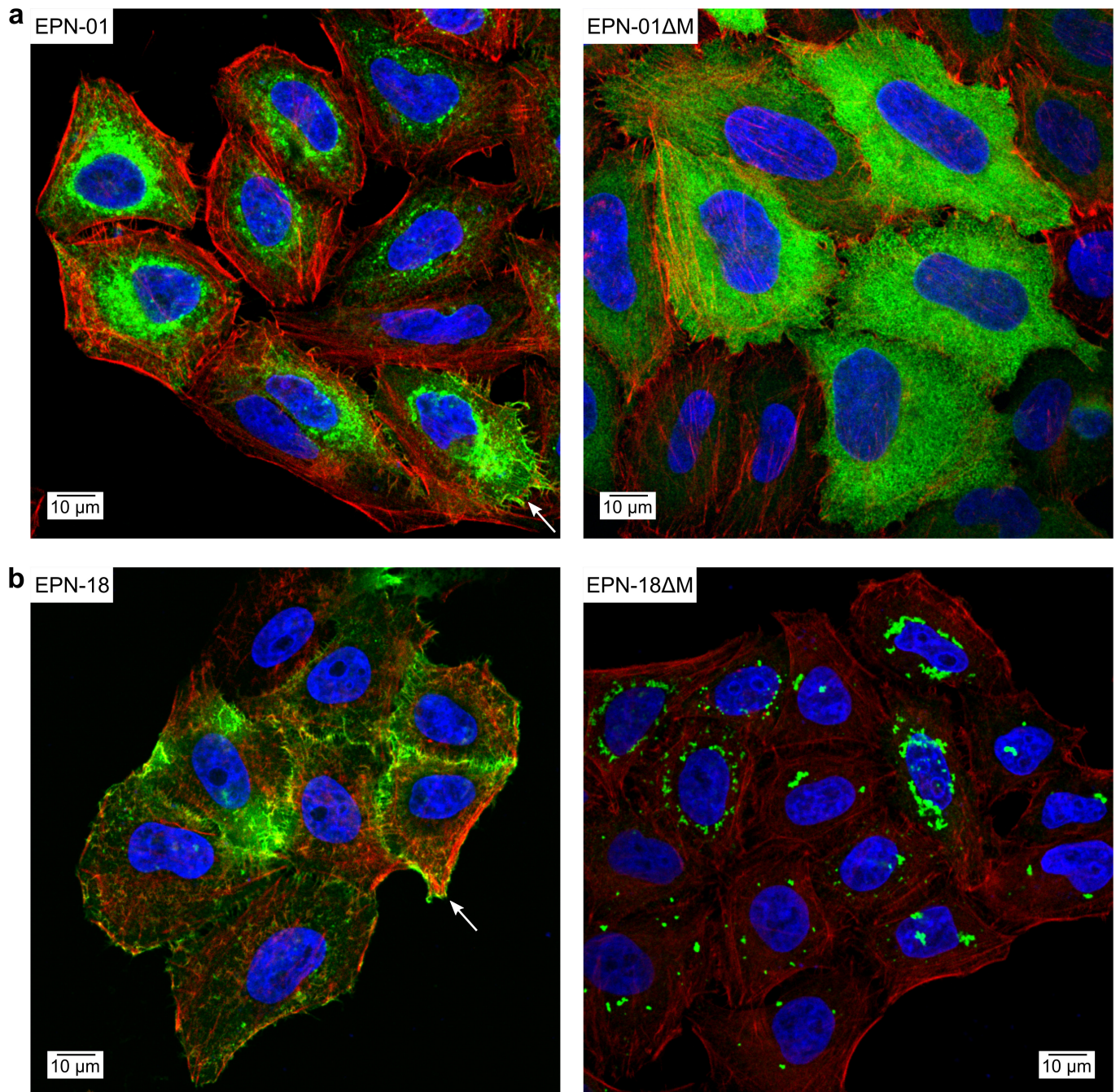
Extended Data Figure 3 | EPN-01 constructs are released within membrane vesicles. **a**, Detergent-dependent antibody accessibility of EPN-01 protein released from mammalian cells. Western blots of EPN-01 harvested from mammalian cell supernatants (top two blots) or non-enveloped EPN-01* nanocages purified from *E. coli* (bottom two blots) are shown. Samples were treated with 0.1% or 0.5% CHAPS, as indicated. Lane 1 shows input protein, and lanes 2–5 show bound and unbound fractions from immunoprecipitations with anti-Myc-agarose (lanes 2 and 3) or control anti-rabbit-IgG-agarose (lanes 4 and 5). **b**, Detergent-dependent aldolase substrate accessibility of EPN-01* protein released from mammalian cells. EPN-01* produced in 293F cells (top) or *E. coli* (bottom) were monitored for KPDG aldolase activity in the presence or absence of detergent. Note that the protein produced in mammalian cells only shows enzymatic activity in the presence of detergent, whereas the

bacterially produced protein is equally active in both conditions. **c**, Size distribution of EPNs released into the culture supernatants of 293T cells expressing EPN-01*. $n = 142$ EPNs total; mean size, 107 ± 44 . **d**, EPN-01* and associated cellular proteins are released from 293T cells. Western blots showing levels of I3-01-Myc, EPN-01*, ALIX and actin expressed in 293T cells (left) and released into the culture supernatant (right). **e**, Numbers of EPN-01* protein nanocages encapsulated within six different EPN-01* vesicles of varying sizes. Nanocage numbers were determined by visual counting of reconstructed EPNs from cryo-tomographic tilt series. An average EPN vesicle of diameter 110 nm has a volume of $7 \times 10^5 \text{ nm}^3$ and each nanocage has an estimated volume of approximately $8 \times 10^3 \text{ nm}^3$. A 110 nm diameter vesicle could therefore theoretically contain as many as 90 close-packed nanocages, but they actually contain, on average, around 14 nanocages.



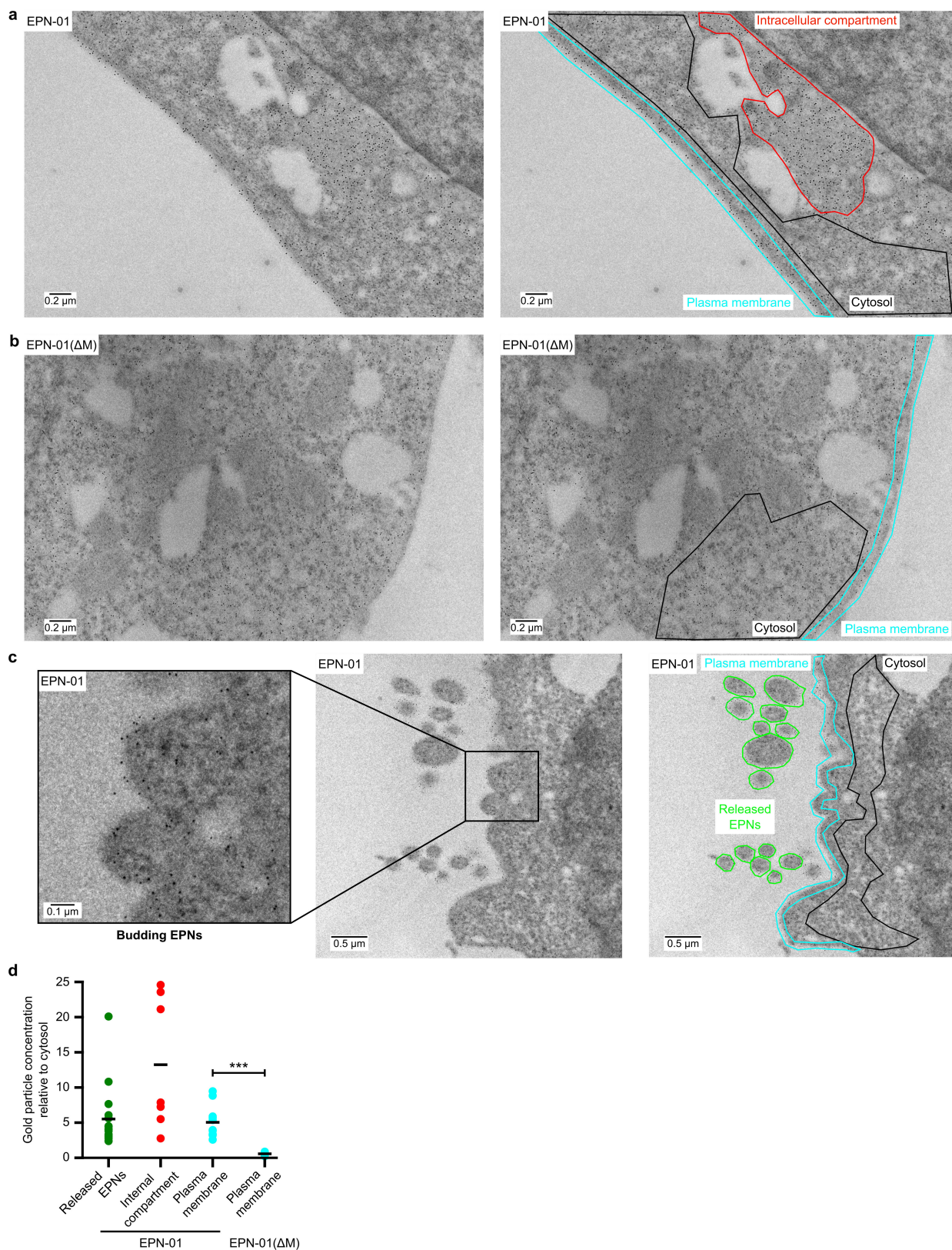
Extended Data Figure 4 | EPN-01* protein released from 293T cells assembles into nanocages that closely correspond to the designed I3-01 structure. **a**, Cryo-EM image of extracellular EPN-01* nanocages released from vesicles by CHAPS detergent treatment. **b**, Cryo-EM 2D class averages showing the ten most prevalent classes, together with the

numbers of particles in each class. **c**, Fourier shell correlations between the two half model charge density maps unmasked (blue curve) and masked (magenta curve). Orange curve shows Fourier shell correlations between the final charge density model and the I3-01 design model (with an additional residue built into the N terminus).



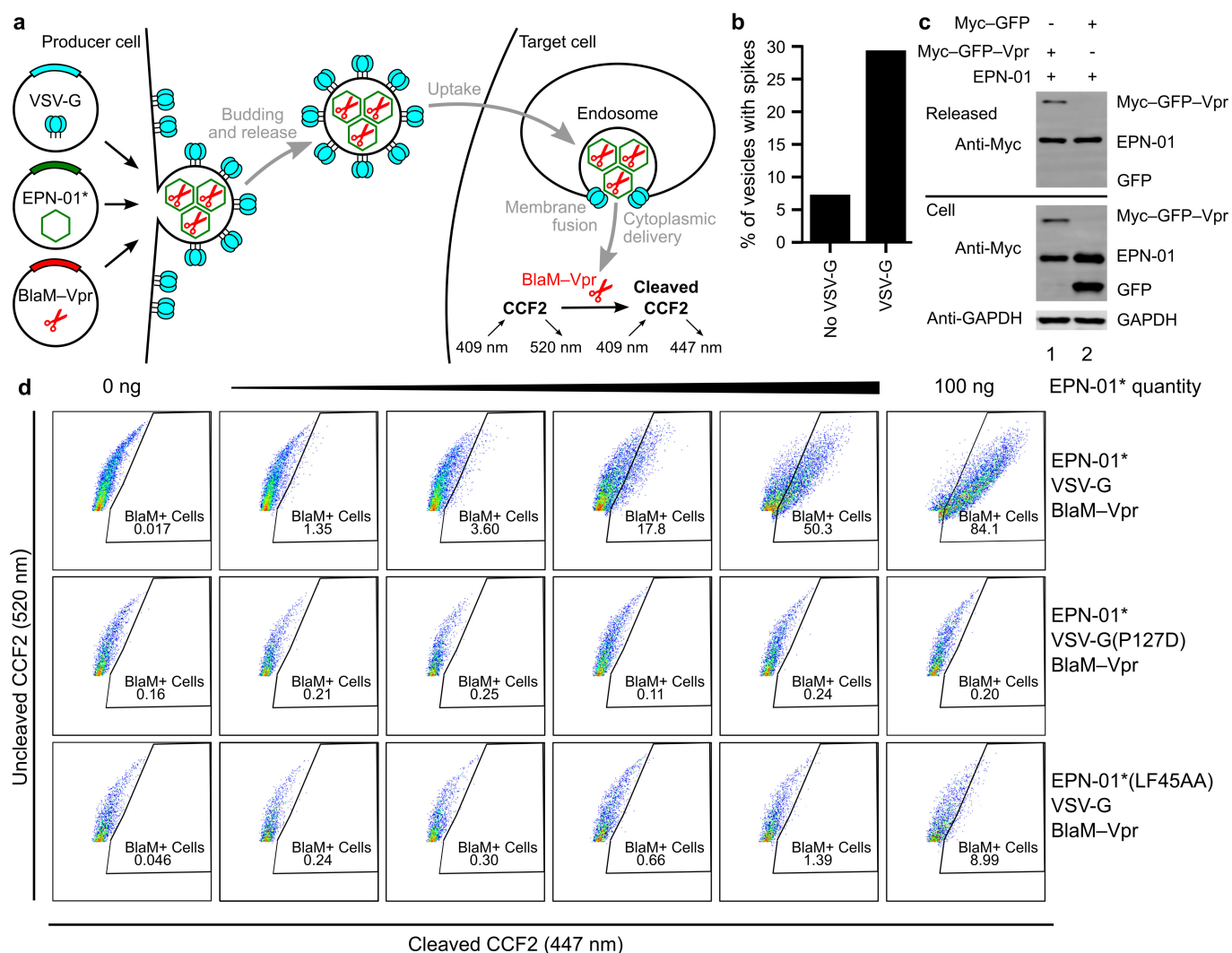
Extended Data Figure 5 | Intracellular localization of EPN-01 and EPN-18. **a**, Confocal fluorescence images of HeLa cells transfected with EPN-01 (left) and EPN-01(Δ M) (right) stained for Myc (green), DNA (blue) and actin (red). Note that EPN-01 is localized primarily in intracellular compartments and also at the plasma membrane (white arrow),

whereas EPN-01(Δ M) is cytoplasmic. **b**, Confocal fluorescence images of HeLa cells transfected with EPN-18 (left) and EPN-18(Δ M) (right). Note that EPN-18 is predominantly localized at the plasma membrane (white arrow), whereas EPN-18(Δ M) concentrates at internal puncta.



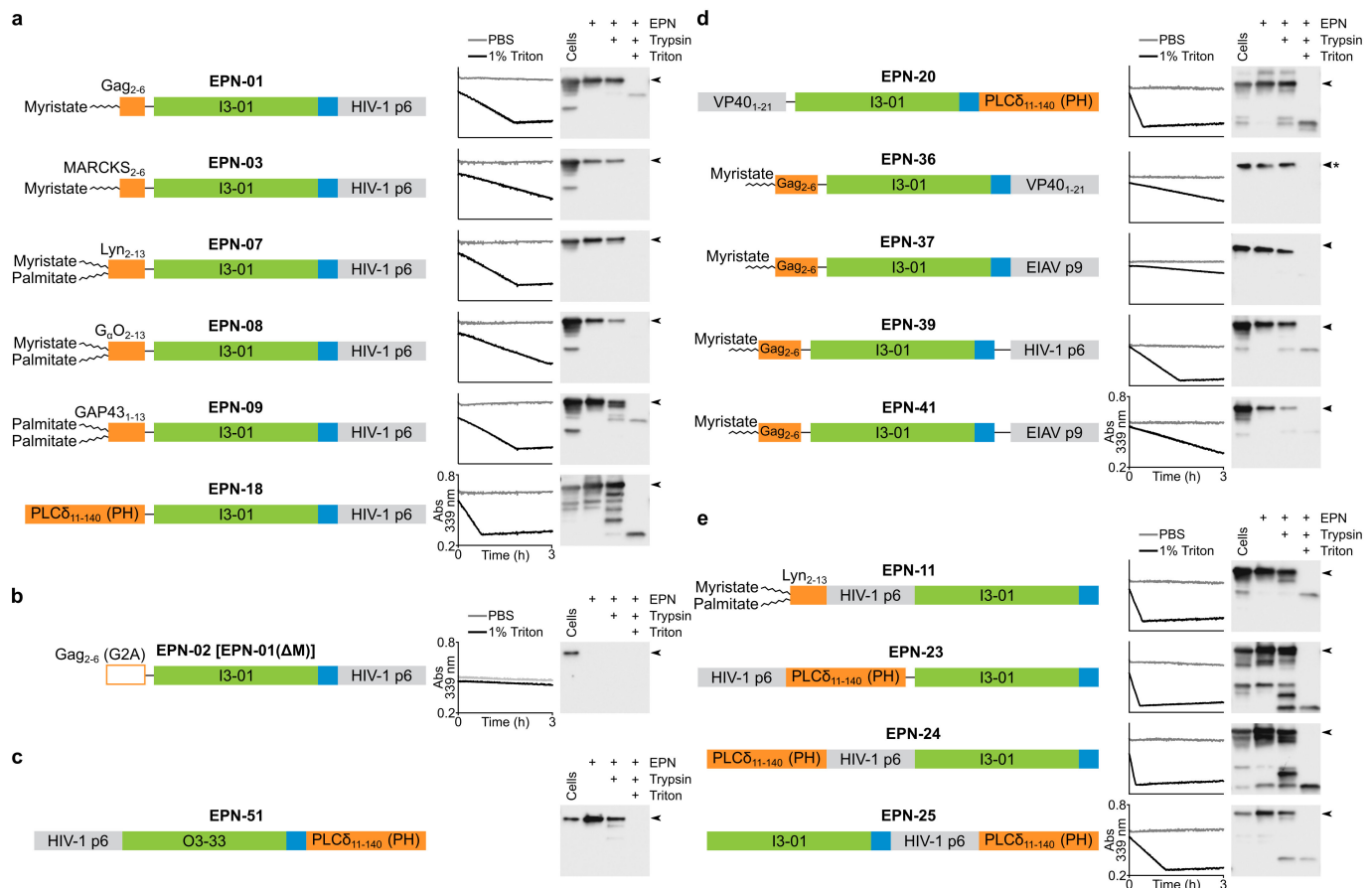
Extended Data Figure 6 | Intracellular localization and extracellular release of EPN-01. **a**, Immunogold labelling of thin sections from 293T cells expressing EPN-01. Right image shows the boundaries of plasma membrane (cyan), cytosolic (black) and intracellular compartments (red) used for quantification (see **d**). **b**, Immunogold labelling of thin sections from 293T cells expressing EPN-01(ΔM). Right image shows the boundaries used for quantification, colour-coded as in **a**, except that the intracellular compartment(s) could not be identified in this case because the EPN-01(ΔM) mutant does not localize there. **c**, Immunogold labelling of thin sections from 293T cells expressing EPN-01 showing budding and

released EPNs. Right image shows the boundaries used for quantification as in **a**, plus boundaries to quantify gold particle densities in released EPNs (green). Image to the left shows an expanded view of two EPNs that appear to be budding from the plasma membrane. **d**, Image quantification confirmed that EPN-01 is enriched at the plasma membrane, within intracellular compartments, and in released EPNs. The *P* value corresponding to the difference between plasma membrane localization of EPN-01 and EPN-01(ΔM) is 7×10^{-5} , indicated by the three asterisks (unpaired *t*-test, $n = 11$ images for EPN-01 and $n = 9$ images for EPN-01(ΔM)).



Extended Data Figure 7 | EPNs can package biological cargoes and deliver them to the cytoplasm of target HeLa cells. **a**, Schematic illustration showing the production, assembly, and release of EPNs incorporating BlaM-Vpr and VSV-G proteins (left), and detection of uptake and target cell membrane fusion using a BlaM colourimetric activity assay (right). **b**, Co-expression of VSV-G increases the number of vesicles that contain spikes, as evaluated by scoring >140 vesicles as either 'containing' or 'not containing' surface spikes (images like the one shown in the inset of Fig. 3b were scored independently by two different people, one blinded, and their counts were averaged). **c**, Western blots showing cellular expression and release of EPN-01 and Myc-tagged GFP constructs

with or without fused Vpr (see Supplementary Table 3 for sequence information). Top blot shows released protein, middle and bottom blots show expression of Myc-tagged proteins and GAPDH in whole-cell lysates, respectively. Lane 1 shows co-expression of EPN-01 with Myc-GFP-Vpr, lane 2 shows co-expression of EPN-01 with Myc-GFP. **d**, Flow cytometric analyses of HeLa cells loaded with the fluorescent CCF2 β -lactamase substrate and incubated with increasing quantities of wild-type EPN-01*/VSV-G/BlaM-Vpr (top row), EPN-01*/VSV-G(P127D) mutant/BlaM-Vpr (middle row), and EPN-01*(LF45AA) mutant/VSV-G/BlaM-Vpr (bottom row).



Extended Data Figure 8 | Aldolase and protease protection assays for EPNs with a variety of functional elements and protein architectures. Schematic illustrations and analyses of the 16 EPN constructs that yielded robust EPN biogenesis are shown, as well as one negative control. Each panel shows the construct, a representative plot of aldolase activity in the presence (black line) and absence (grey line) of detergent, and a western blot analysis of the protease protection assay. Arrowheads next to each blot denote the full-length protein. Aldolase activity was monitored by disappearance of absorbance at 339 nm. **a**, Different membrane-binding elements support EPN formation. **b**, EPN-02, also referred to as EPN-01(ΔM), is a negative control construct in which the myristoylation site was inactivated by mutation. Both assays reveal that EPN-02 protein was

not released from cells. **c**, EPN-51, which uses the designed 24-subunit protein assembly O3-33 as a self-assembly domain, forms an EPN with an intact membrane envelope. The aldolase assay was not included because O3-33 is not an aldolase. **d**, Different ESCRT-recruiting elements can support EPN formation. The asterisk next to the blot of EPN-36 signifies that the blot was overexposed: EPN-36 reproducibly yielded fainter bands on western blots than would be expected based on its aldolase activity and analyses of SDS-PAGE gels stained with Coomassie. **e**, Membrane-binding, self-assembly, and ESCRT-recruiting elements can function from different positions within EPN constructs. EPN-11 is a permutation of EPN-07, while EPN-23, EPN-24, and EPN-25 are permutations of EPN-18.

Capturing pairwise and multi-way chromosomal conformations using chromosomal walks

Pedro Olivares-Chauvet¹, Zohar Mukamel¹, Aviezer Lifshitz¹, Omer Schwartzman¹, Noa Oded Elkayam¹, Yaniv Lubling¹, Gintaras Deikus², Robert P. Sebra² & Amos Tanay¹

Chromosomes are folded into highly compacted structures to accommodate physical constraints within nuclei and to regulate access to genomic information^{1,2}. Recently, global mapping of pairwise contacts showed that loops anchoring topological domains (TADs) are highly conserved between cell types and species^{3–8}. Whether pairwise loops^{9–14} synergize to form higher-order structures is still unclear. Here we develop a conformation capture assay to study higher-order organization using chromosomal walks (C-walks) that link multiple genomic loci together into proximity chains in human and mouse cells. This approach captures chromosomal structure at varying scales. Inter-chromosomal contacts constitute only 7–10% of the pairs and are restricted by interfacing TADs. About half of the C-walks stay within one chromosome, and almost half of those are restricted to intra-TAD spaces. C-walks that couple 2–4 TADs indicate stochastic associations between transcriptionally active, early replicating loci. Targeted analysis of thousands of 3-walks anchored at highly expressed genes support pairwise, rather than hub-like, chromosomal topology at active loci. Polycomb-repressed *Hox* domains are shown by the same approach to enrich for synergistic hubs. Together, the data indicate that chromosomal territories, TADs, and intra-TAD loops are primarily driven by nested, possibly dynamic, pairwise contacts.

When chromatin is fixed, digested with frequent cutters (such as DpnII) and re-ligated, high-molecular-weight DNA products can be purified, representing concatenation of 30–60 DNA fragments with a total length of 10,000–20,000 base pairs (bp) (Fig. 1a). Such products are likely to represent spatial proximity between multiple loci that were captured at the time of fixation. Nevertheless, the multi-way proximity relations emerging from such products have so far been studied primarily by profiling pairwise ligation contacts, using techniques such as circularized chromosome conformation capture (4C), chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) or Hi-C. To enable a more comprehensive analysis of multi-way chromosomal proximities, we developed a technique involving the generation and selection of high-molecular-weight chromosome conformation capture (3C) DNA, serial dilution and distribution of the resultant material into 96-well plates (1 pg into each well), Φ 29-DNA-polymerase-based amplification, sonication and labelling with well-specific barcodes. This was followed by amplification, sequencing and computational assembly of chains of ligation junctions that are defined here as ‘C-walks’ (Fig. 1b and Methods).

We generated C-walk libraries from human K562 cancer cells and from mouse embryonic stem (mES) cells, which are both highly established model systems for the exploration of chromosome conformations. Analysis of the inferred C-walk contacts indicated that pairwise genomic distance distributions and inter-chromosomal rates are similar to those observed in standard Hi-C (Supplementary Table 1 and Extended Data Fig. 1a, b). Analysis of the C-walk coverage and

size distribution (Fig. 1c) indicated that our approach (in K562 cells) assembled 48% of the pairwise ligations into C-walks of size 4 or more, and 14% to C-walks of size 8 or more. We recovered 45,200 fragments participating in C-walks involving 16 or more fragments. To confirm the accuracy of our reconstruction approach, we applied the procedure to linear DNA (Extended Data Fig. 1c, d). We estimate that the complete assay associated fragments into C-walks with over 99% accuracy (Extended Data Fig. 1e, f), limiting the potential effect of amplification, labelling and sequencing errors on the downstream statistical analysis of C-walk distributions.

We calculated the average rate of Hi-C pairwise inter-chromosomal contacts to be 22% in K562 and 17% in mES cells. Higher-order analysis showed, however, that distinct classes of proximity ligation events influence this rate. We observed walks visiting only one chromosome (class I; Fig. 1d) and walks linking two chromosomes through one or more contacts (class II), but also walks bringing together fragments from three or more chromosomes (class III). While class III C-walks may theoretically represent true multi-chromosomal hubs, their internal structure (namely, their lack of intra-chromosomal hops) and nearly uniform distribution of pairwise contacts (Extended Data Fig. 1g, h) suggest that these are products resulting from spurious ligations and possibly explaining quality differences between in-solution and in-nucleus 3C protocols¹⁵. After filtering out class III C-walks, we estimated the pairwise inter-chromosomal Hi-C contact rate to be 7–10% (Fig. 1e, f). Interestingly, more than half of the valid inter-chromosomal (class II) C-walks (for $C > 4$) link two chromosomes through more than one contact. Classification of the resultant inter-chromosome interfaces suggests that these are strongly restricted by the topological domain structure in each of the contacting chromosomes (Extended Data Figs 2a–f and 3a–e).

Each class I C-walk explores a chromosomal territory by a series of hops. The data show that, as suggested previously^{6,7}, the pairwise chromosomal distance bridged by a class-I hop is governed by a power-law-like regime: the aggregated probability of ‘big’ (for example, 1–100 Mb) hops is similar in scale to the aggregated probability of making ‘small’ (for example, 10 kb to 1 Mb) hops (Extended Data Fig. 1b). On the basis of this rule, the probability of a walk remaining constrained within a smaller fraction of the chromosome should decrease exponentially with the number of hops. Nevertheless, analysis of specific regions (Fig. 2a) suggested that a notable fraction of the class I C-walks are restricted to less than 1 Mb, and are frequently fully contained within one TAD (Fig. 2b). Moreover, C-walks that visit more than one TAD typically link elements at high chromosomal distances (Fig. 2c).

To quantify these observations, we defined the span of a class I C-walk to be the chromosomal distance between the minimal and maximal chromosomal coordinates visited by the walk. We also classified C-walks based on their compartment (active or inactive, using only the first TAD visited to avoid indirect correlations). The spans for increasingly long C-walks (Fig. 2d and Extended Data Fig. 3f) are

¹Department of Computer Science and Applied Mathematics and Department of Biological Regulation, Weizmann Institute, Rehovot 76100, Israel. ²Icahn Institute and Dept. of Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York 10029, USA.

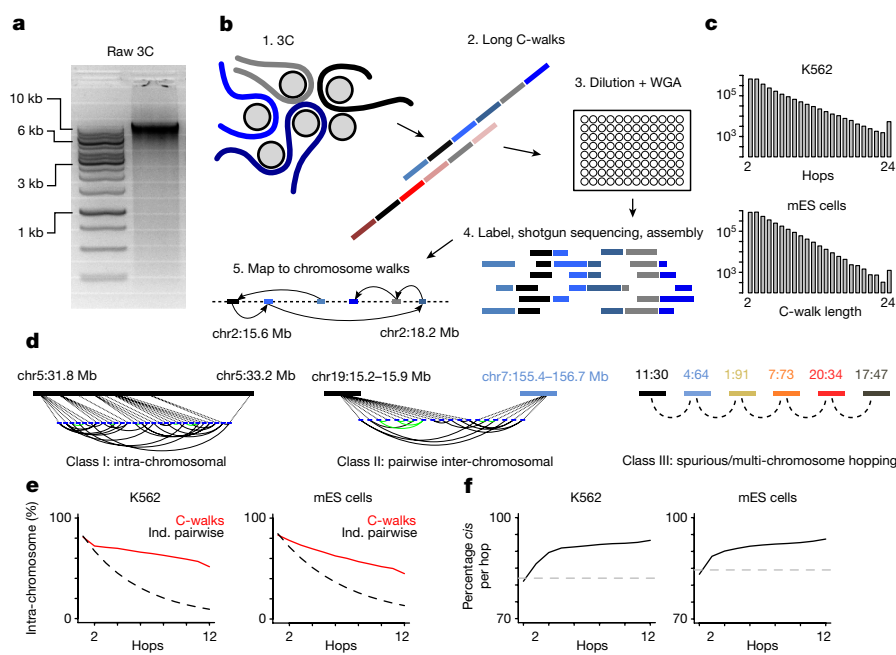


Figure 1 | C-walk mapping. **a**, Gels showing typical 3C DNA fragment length distributions after digestion and ligation of cross-linked nuclei and DNA purification. **b**, C-walk shotgun sequencing protocol schematics. WGA, whole genome amplification. **c**, Number of contacts recovered (y axis) within C-walks of length (x axis) 2 to 24. **d**, Example of three C-walk classes. Graphs depict C-walk hops as arcs (bottom) and the mapping of C-walk loci to the linear chromosome (top, black box). Green arcs are hops within 5 kb. Loci covered by class III C-walks are labelled as chromosome (chr):coordinate (in Mb). **e**, The fraction of C-walks that remain within a single chromosome is stratified by the number of hops (red curve). The expected fraction assuming independent (ind.) inter-chromosomal hops is shown as a dashed line. **f**, The rate of inter-chromosomal hopping is shown, using only C-walks class I and II, and stratifying by C-walk length. We note that detecting spurious C-walks (class III) is impossible for $C = 2$ and is imperfect for $C = 3$, but for $C > 3$ the estimated inter-chromosomal hop rate converges, suggesting that filtering is robust.

shown to be restricted to 1 Mb or 10 Mb for a large fraction of the cases (also validated using direct PacBio sequencing; Extended Data Fig. 3g). The overall trend (Fig. 2e) suggests that 40% of the long C-walks visiting active regions and 20% of those visiting inactive regions are constrained within less than 1 Mb (Extended Data Fig. 3h; $P < 10^{-3}$). The number of TADs visited by long C-walks (Fig. 2f) reflected a similar behaviour, but here the difference between active and (overall larger) inactive TADs was less notable, suggesting that TAD structure, more than the linear span of the chromosome, restricts the dimensions of C-walks. The data therefore suggest that TADs organize chromosomal

territories by forming intra-TAD spaces in which C-walks can be fully contained, and inter-TAD interfaces connecting these spaces.

The topology within TADs can be interrogated by studying intra-TAD C-walks. We normalized the span of such C-walks by their TAD size, and examined separately for active and inactive TADs how quickly a C-walk would span at least half of the TAD space (in terms of the number of hops) (Fig. 2g). We observed that hopping within active TADs saturates the domain more slowly than within inactive TADs, even when controlling for TAD size ($P < 10^{-3}$; Extended Data Fig. 4a, b). Moreover, when TADs become larger (for example, from 1–1.5 Mb

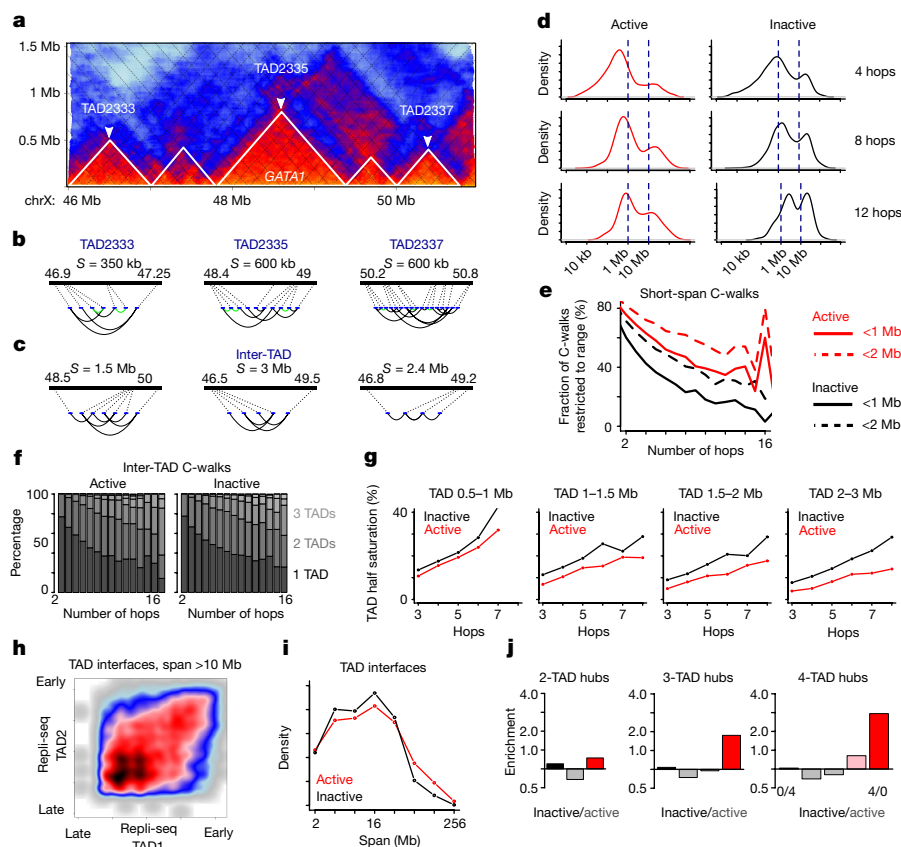


Figure 2 | Intra-chromosomal hierarchical topology, in and between TADs. **a**, A Hi-C map shows the TAD structure in the 5 Mb around the *GATA1* locus in K562. **b**, Examples for intra-TAD C-walks. Arcs at the bottom depict hops, linked loci are mapped to a linear scale at the top. The span (S) of each C-walk is shown. **c**, Examples for inter-TAD C-walks. **d**, Distribution of C-walk span stratified by number of hops. C-walks are classified into active and inactive based on the TAD visited (Methods). **e**, Shown are the fractions of C-walks restricted to a span of less than 1 Mb (solid line) or 2 Mb (dashed line), as a function of the number of hops. **f**, Shown are the numbers of distinct TADs visited by a C-walk, stratified according to the number of hops. **g**, The fraction of intra-TAD C-walks with a span that covers at least half of the TAD containing them, as a function of the number of hops (x axis) and TAD type (colour). Data are shown separately for C-walks within TADs for four different size ranges. **h**, Heatmap shows the time of replication of TADs that are linked by intra-chromosomal C-walks with at least a 10 Mb span. **i**, Distribution of distances between TADs that are linked by C-walks, shown separately for pairs of active and inactive domains. **j**, For each intra-chromosomal C-walk we counted the number of active and inactive domains visited. Shown are fold change enrichment (observed number over expected number) for C-walks visiting only inactive (0/4, black), only active (4/0, red), or mixed (grey, pink) TADs.

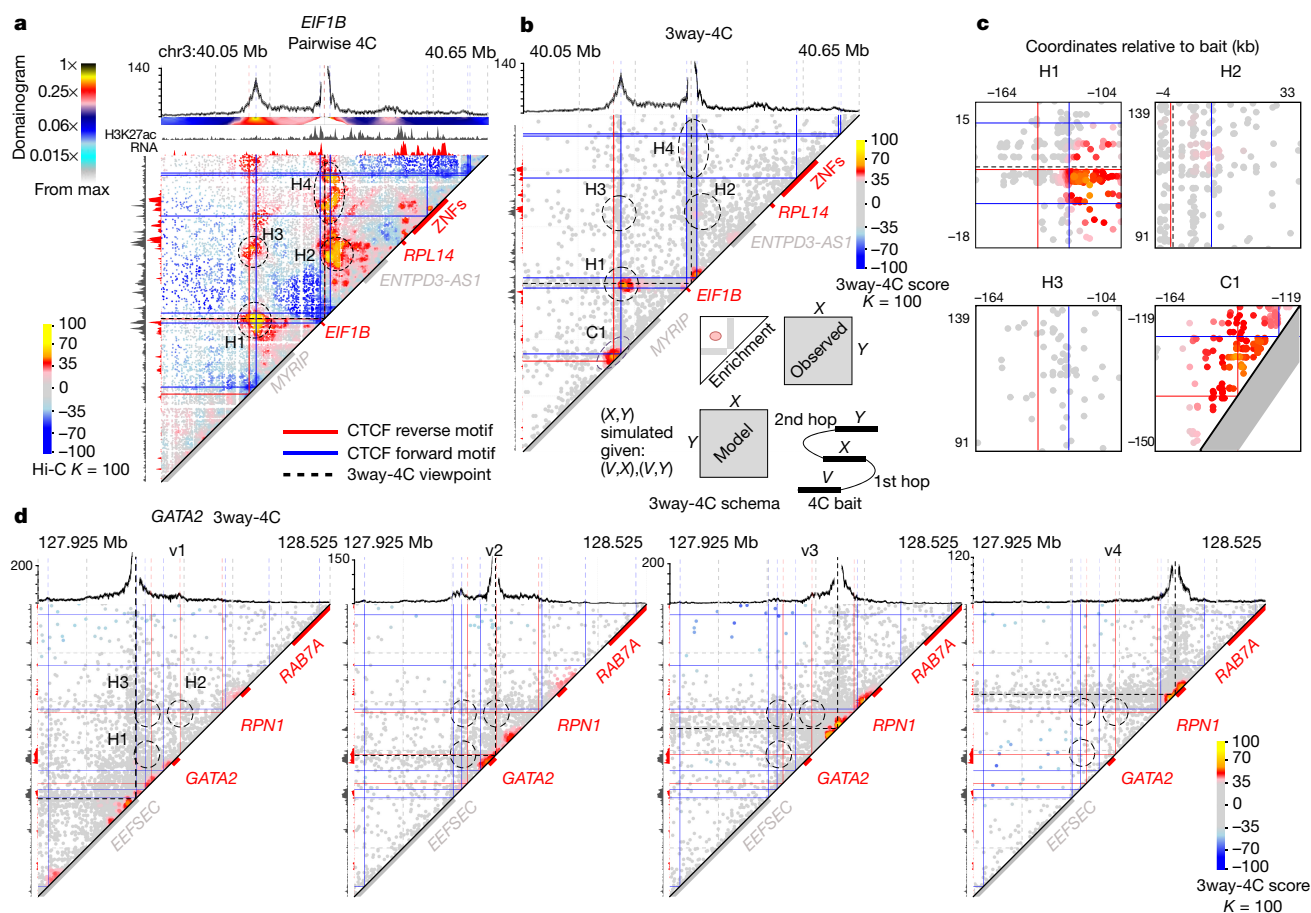


Figure 3 | Hubs exist but are not enriched at highly expressed genes.

a, Shown are pairwise contact enrichments around the *EIF1B* locus. Each point represents a contact, colour-coded according to the density of Hi-C contacts around it, normalized by a probabilistic model assuming total Hi-C coverage per element and a global regime defining contact probabilities per genomic distance (Methods). Key annotation is indicated on the diagonal, marking highly active genes in red and the rest in grey. CTCF loci form grids coloured according to motif orientation (blue for forward motif and red for reverse motif). One-dimensional epigenomic information is shown on the sides. A high-resolution 4C profile and domainogram highlighting contacts from a viewpoint next to the *EIF1B*

promoter is shown on top. **b**, 3way-4C generates thousands of 3-walks $V-X-Y$ originating from a viewpoint V (bottom right schematics). We compared the observed distribution of pairs (X,Y) that contact V to a model that takes into account the pairwise contact distributions of V , X and Y . The top colour-coded triangular heatmap depicts the enrichment scores computed for each triplet $V-X-Y$. The top profile shows the 4C pairwise profile for reference. Key regions of interest are marked H1–4 and C1. **c**, Enlarged maps of 3way-4C contacts at pairwise hotspots (H1–3) and the contact cluster C1. **d**, 3way-4C data are shown for four baits around the *GATA2* gene (see Extended Data Fig. 6a for a Hi-C map of the locus).

to 1.5–2 Mb), the saturation is not delayed (in terms of the number of hops) significantly. C-walks can also be used to evaluate the topology of intra-chromosomal TAD–TAD interfaces. Figure 2h indicates that TAD–TAD interfaces bring together domains with a correlated time of replication, and Fig. 2i shows that the linear distance span by such interfaces is larger for active TADs ($P < 10^{-3}$). Furthermore, C-walks visiting two, three or four TADs within the same chromosome are up to threefold enriched for linking multiple active TADs (Fig. 2j; $P < 10^{-3}$). Similar trends are observed in mES cells C-walk data (Extended Data Fig. 4c–g). In summary, active and inactive TADs form mini-territories that represent building blocks for chromosomal territories, but active TADs show a distinct internal topology consistent with a less compact organization, enrichment of long-range intra-chromosomal contacts with other active TADs, and potentially multi-TAD hubs.

C-walk analysis suggested that long-range contacts generally tend to form pairwise and not multi-way interfaces (Extended Data Fig. 4h–j). We next wished to examine this question at higher resolution and to test whether contacts linking CTCF/cohesin-binding sites, promoters or enhancers, form nested loops within loops or can be detected in the context of multi-way hubs. We therefore developed a multiplexed 3way-4C procedure¹⁶ that allowed us to profile up to 50,000 3-walks across panels of designed viewpoints (Supplementary Table 2). Each

3-walk determines the association of two loci to one viewpoint. If we assume a viewpoint (V) synergistically contacts two loci, X and Y , then a large (that is, larger than otherwise expected) fraction of the 3-walks where V contacts X should also visit Y . On the other hand, if contacts of V and X are enriched independently of contacts between V and Y , then the probability of observing the triplet (V,X,Y) should be predictable by the probability of observing the pairwise contacts (V,X) , (V,Y) and (X,Y) . We note that once we assume a contact between V and X exists, the background probability model must take into account sampling the next hop Y from the (pairwise) chromosomal neighbourhoods of either X or V (Extended Data Fig. 5a). We developed a model that compares the distribution of 3-walks observed in the data to the distribution expected from pairwise contacts, allowing screening for hubs that are significantly enriched for tri-way contacts (Extended Data Fig. 5b–f). 3way-4C analysis distinguishes between complex Hi-C contact maps that can be explained by independent mixtures of pairwise contacts and maps that contain true synergistic contact hubs.

We first applied the targeted 3way-4C analysis to characterize the conformation around several highly expressed genes in K562 cells. Hi-C analysis around the highly expressed *EIF1B* locus (Fig. 3a) indicated two strong loops (H1 and H2) enclosing two TADs that flank the gene. Interestingly, the Hi-C results indicated that a weaker loop

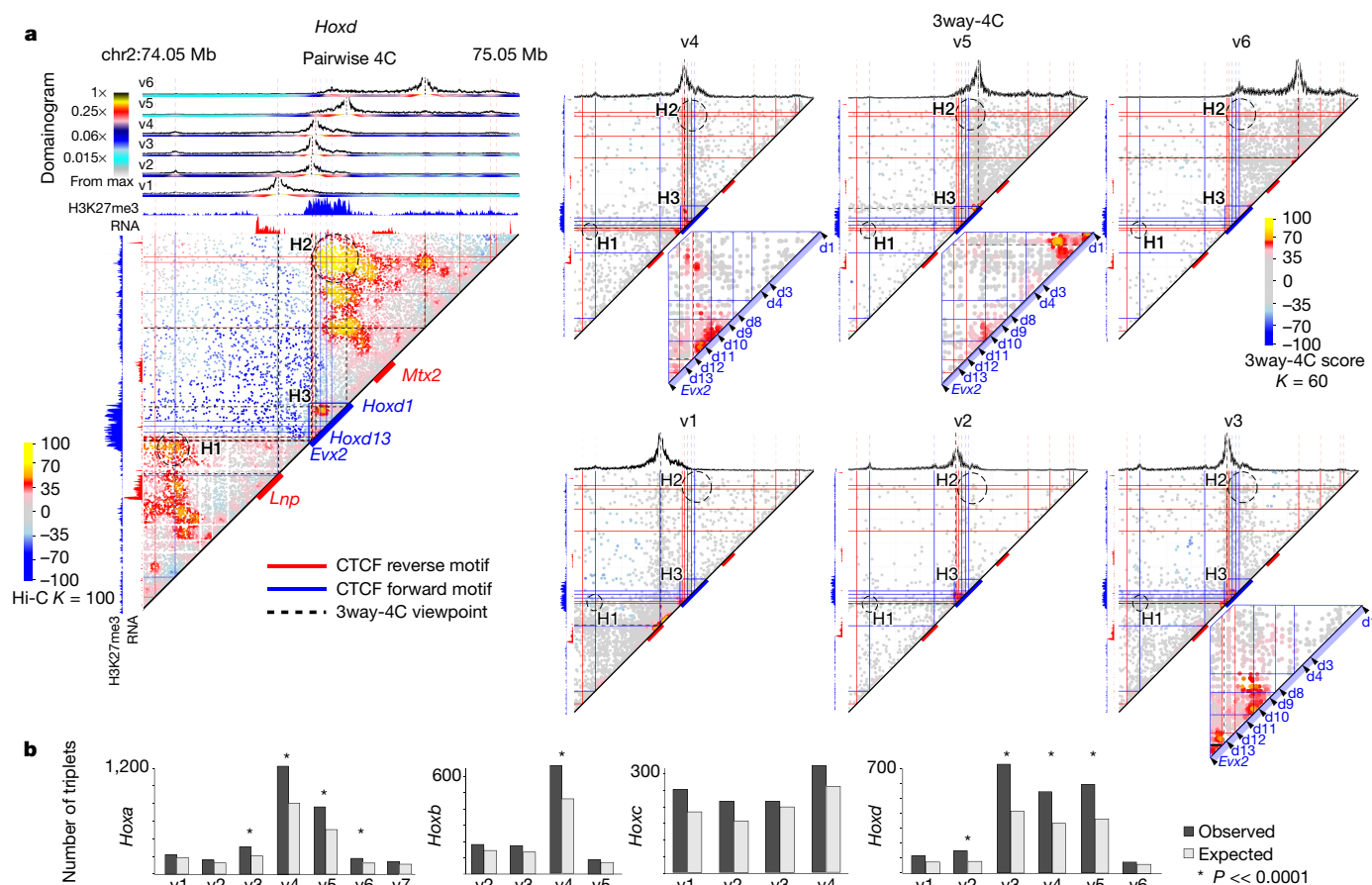


Figure 4 | Synergistic contacts at the *Hoxd* Polycomb domain. **a**, Similar to Fig. 3, showing data on pairwise and 3way-4C contacts around the *Hox* clusters in mES cells. Insets show a magnified view of the Polycomb-enriched *Hoxd* TAD. d1, d3, and so on, indicate *Hoxd* gene numbers.

b, Bar graphs show observed and expected triplet contacts within the four *Hox* domains, as assayed from 21 different viewpoints. * $P < 0.0001$, chi-squared test.

(H3) merges the two TADs together, while another broad loop (H4) seems to combine H2 with a third TAD containing the highly expressed *RPL14* gene. Using 3way-4C we generated 102,189 pairwise contacts, and 15,638 3-walks anchored on a viewpoint near the *EIF1B* promoter. Modelling these contacts (Fig. 3b) showed that triplet distributions are predictable from the pairwise profile in almost all cases (grey points). We detected some synergistic contacts ‘within’ the H1 loop (left of the *EIF1B* promoter, right of the dual CTCF site marked by the red and blue line) (Fig. 3c; $P < 10^{-3}$). Importantly, we observed no synergistic contacts closing the two loops at H3 (Fig. 3c). On the other hand, we observed hubs linking the promoter element within a mini-loop at the base of H1 (denoted C1; Fig. 3c; $P < 10^{-3}$). Similar observations were made using data on the *GATA2* (Fig. 3d and Extended Data Fig. 6a), *GATA1*, *FTL* and *ANK1* active loci (Extended Data Figs 6 and 7). A somewhat more complex, but still non hub-like triplet distribution was observed in the β -globin locus (Extended Data Fig. 8). In summary, 3way-4C analysis indicates that long-range hubs associating CTCF loops or enhancers at active genes can be observed, but at frequencies that follow a model of independent pairwise contacts and not synergistic, cooperative contact hubs.

Next, we used our strategy to screen for higher-order structure at *Hox* gene clusters in mES cells. These key domains are targeted for repression by the Polycomb machinery and serve as a paradigm for TAD-supported gene regulation¹⁷. We generated 3way-4C profiles from 23 viewpoints located in and around the four *Hox* clusters (Supplementary Table 2). Using our high-resolution Hi-C analysis approach we detected a rich looping structure that packages the *Hox* clusters into H3K27me3-enriched domains. As illustrated in Fig. 4a, the *Hoxd* locus is flanked

by two strong loops (H1 and H2), which in turn delimit the two TADs encompassing the expressed genes *Lnp* (also known as *Lnpk1*) and *Mtx2*. Remarkably, the *Hoxd* module is organized into a third smaller loop (H3). This domain (Fig. 4a, blue frames) is organized by an internal array of CTCF sites positioned between *Evx2*–*Hoxd13* to *Hoxd8*, and a second region lacking CTCF sites, which together form contacts across the entire *Hox* domain through a strong non-CTCF anchored loop. 3way-4C analysis on six viewpoints (v1–v6, Fig. 4a), indicate weak or no enrichment of synergistic *Hoxd* contacts from within the H1 TAD (viewpoint v1) or the H2 TAD (viewpoint v6). On the other hand, viewpoints within the H3K27me3 domain (v2–v5) show stronger evidence for synergistic enrichment of triplets on the H3 loop. The Polycomb domain structure in the *Hoxa* cluster is similar, but is further divided into two loops (Extended Data Fig. 9), which are both synergistically observed from viewpoints within the domain, but not outside it. A similar, though more complex structure is observed in the *Hoxc* and *Hoxb* loci (Extended Data Fig. 10; a direct statistical test for triplet enrichment in the Polycomb domains is shown in Fig. 4b). In summary, the *Hox* clusters show a distinct topological architecture with overlapping nested TADs and a Polycomb-TAD, in which the Polycomb-TAD shows evidence for synergistic multi-way contacts. This indicates that the mechanisms packaging *Hox* clusters into Polycomb domains involve more than pairwise (CTCF-mediated) looping, and provide a good test case for analysis of chromatin hubs in repressive contexts.

We studied chromosomal topologies using C-walks: series of hops between chromosomal positions that reflect a shared three-dimensional proximity. We used C-walks to address some questions regarding the hierarchical organization of chromosomal topologies. Our results show

that chromosomal territories are more strongly defined than previously appreciated by Hi-C, that the interface between territories is likely to occur at the level of TADs⁴, and that within territories, TADs define a mini-territory such that C-walks can be fully contained within a TAD even after a large number of hops. We used these observations in conjunction with a targeted 3way-4C approach to study higher-order looping structures at highly transcribed loci in active TADs. We found that while these loci are organized by strong pairwise loops, we could not detect hubs that bring together synergistically more than two loci. When we turned our attention to repressed Polycomb domains in mES cells, we did find evidence for synergistic clustering of *Hox* genes. This suggests that our assays for multi-way contact hub detection provides sufficient power to detect hubs if these occur at frequencies that are at the scale of pairwise contacts. However, our data do not necessarily preclude such structures existing transiently or in contexts we did not analyse in depth.

Chromosomes must be folded efficiently to allow both compactness and functionality. Functionality must be understood while considering dynamic processes such as DNA replication and transcription. Pairwise looping can facilitate packaging, and has recently been suggested to be driven by CTCF (providing binding site specificity) and cohesin, possibly by a dynamic process of loop extrusion that is blocked by properly oriented CTCF binding sites^{18,19}. Our data indicate that nested topologies allow accommodation of most regulatory interactions within TADs that actively transcribe genes, such that violations of the pairwise structure around promoters and enhancers are mostly indirect or perhaps transient. When seemingly higher-order conformation is observed on typical Hi-C contact maps, our data suggest that a mixture of nested structures can explain it in many cases. It remains to be seen how general this rule is. For example, in the context of gene repression, aggregation of chromosomal elements may involve foci of more than two elements, as suggested by data on Polycomb domains in this and other studies^{20–23} or by data on X chromosome inactivation²⁴. Furthermore, if key regulatory contacts violate the nested structure in a highly transient fashion, their footprint on contact maps may be much smaller than their functional impact, resulting in architectures that cannot be detected by our current methods. More experimental data on pairwise and higher-order chromosomal conformation²⁵, together with quantitative modelling of the dynamics that may explain such data^{26,27}, are needed to further strengthen our understanding of how higher-order chromosomal structures participate in genome regulation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 December 2015; accepted 13 October 2016.

Published online 30 November 2016.

1. Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect. Biol.* **2**, a003889 (2010).
2. Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature Rev. Genet.* **14**, 390–403 (2013).
3. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
4. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).

5. Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Reports* **10**, 1297–1309 (2015).
6. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
7. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
8. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
9. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl Acad. Sci. USA* **111**, 996–1001 (2014).
10. Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* **32**, 3119–3129 (2013).
11. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
12. Zhang, Y. *et al.* Chromatin connectivity maps reveal dynamic promoter–enhancer long-range associations. *Nature* **504**, 306–310 (2013).
13. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
14. Tang, Z. *et al.* CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611–1627 (2015).
15. Nagano, T. *et al.* Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biol.* **16**, 175 (2015).
16. Ay, F. *et al.* Identifying multi-locus chromatin contacts in human cells using tethered multiple 3C. *BMC Genomics* **16**, 121 (2015).
17. Andrey, G. *et al.* A switch between topological domains underlies *HoxD* genes collinearity in mouse limbs. *Science* **340**, 1234167 (2013).
18. Fudenberg, G. *et al.* Formation of chromosomal domains by loop extrusion. *Cell Reports* **15**, 2038–2049 (2016).
19. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112**, E6456–E6465 (2015).
20. Cheutin, T. & Cavalli, G. Polycomb silencing: from linear chromatin domains to 3D chromosome folding. *Curr. Opin. Genet. Dev.* **25**, 30–37 (2014).
21. Schuettengruber, B. *et al.* Cooperativity, specificity, and evolutionary stability of Polycomb targeting in *Drosophila*. *Cell Reports* **9**, 219–233 (2014).
22. Fabre, P. J. *et al.* Nanoscale spatial organization of the *HoxD* gene cluster in distinct transcriptional states. *Proc. Natl Acad. Sci. USA* **112**, 13964–13969 (2015).
23. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).
24. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
25. Williamson, I. *et al.* Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev.* **28**, 2778–2791 (2014).
26. Imakaev, M. V., Fudenberg, G. & Mirny, L. A. Modeling chromosomes: Beyond pretty pictures. *FEBS Lett.* **589**, 3031–3036 (2015).
27. Giorgetti, L. *et al.* Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157**, 950–963 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge E. Yaffe for help with early stages of the experiments, N. Mor for help with mES cells, and members of the A.T. group for discussions. Research at the A.T. group was supported by the European Research Council (EVOEPIC), Flight Attendant Medical Research Institute (FAMRI), and the Israel Science Foundation (ISF). A.T. is a Kimmel investigator.

Author Contributions P.O.-C. and A.T. designed the study with help from Z.M. P.O.-C. and Z.M. developed and optimized the experimental approach with help from N.O.E., O.S. and Y.L. P.O.-C. and A.T. developed the algorithmic approaches and analysed the data, with help from A.L., Y.L. and O.S. G.D. and R.B.P. performed PacBio sequencing. P.O.-C. and A.T. wrote the paper.

Author Information Data have been deposited in the Gene Expression Omnibus under accession number GSE77553. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.T. (amos.tanay@weizmann.ac.il).

METHODS

Cell culture. Human leukaemia cell lines K562 (ATCC CCL-243) were grown in RPMI (GIBCO) supplemented with 10% FBS, 1% glutamine (GIBCO), penicillin (100 U/ml) and streptomycin (100 µg/ml). Murine naive pluripotent cells were grown as previously described²⁸ on gelatin-coated plates. Briefly, mES cells were grown in serum-free chemically defined N2B27-based media: 500 ml KO-DMEM (Invitrogen), 15% knockout serum replacement (Invitrogen; 10828), 5 ml N2 supplement (Invitrogen; 17502048), 5 ml B27 supplement (Invitrogen; 17504044), 1 mM glutamine (Invitrogen), 1% non-essential amino acids (Invitrogen), 0.1 mM β-mercaptoethanol (Sigma), 1% penicillin–streptomycin (Invitrogen), 5 mg/ml BSA (Sigma). Naive conditions for mES cells included 5 µg recombinant human LIF (Peprotech) in addition to 2i: small-molecule inhibitors CHIR99021 (CH, 3 µM; Axon Medchem) and PD0325901 (PD, 1 µM; Axon Medchem). All cell lines were grown at 37 °C and 5% CO₂. Mycoplasma tests were routinely performed on all cell lines. No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

3C. Between 25×10^7 and 50×10^7 cells were used to perform in-nucleus ligation 3C as previously described¹⁵ with a few modifications, in particular the biotin fill-in step was omitted⁷. Briefly, mES were incubated in the presence of trypsin (Biological Industries 03-053-1) for up to 5–8 min at 37 °C and passed through a cell strainer after harvesting to ensure single-cell suspension. Cells (including K562 from now on) were washed once with PBS, resuspended in PBS/FBS 10% and cross-linked in formaldehyde (final concentration of 2%) for 5 min at room temperature. After quenching with 0.125 M glycine and washing with PBS, cells were incubated in permeabilization buffer (10 mM Tris–HCl pH 8, 10 mM NaCl, 0.2% NP-40 alternative (CALBIOCHEM 492016), protease inhibitor cocktail (P8340 Sigma) 10 µl/ml) for 60 min at 4 °C while slowly mixing. Samples were split into 5 M cell aliquots and washed twice with cold PBS. After the second wash, cell pellets were flash frozen on liquid nitrogen.

Pellets were resuspended in 175.5 µl H₂O and 24.5 µl DpnII RE buffer 10× and incubated for 60 min at 37 °C while agitating (750 r.p.m.) in the presence of SDS (final concentration 0.3%). Triton X-100 20% was added to a final concentration of 1.8% and incubated for 60 min at 37 °C while agitating (750 r.p.m.). Aliquots of 2–5 M cells were then digested overnight with 15 µl DpnII (50,000 U/ml, NEB R0543M). Cells were centrifuged for 5 min at 600g and washed with PBS. Incubation of pellets for 20 min at 65 °C was performed to heat-inactivate any residual DpnII. In-nucleus ligation was then performed overnight by incubating nuclei in ligation mix at 16 °C (415 µl H₂O, 5 µl BSA (10 mg/ml NEB) and 5 µl T4 ligation enzyme NEB M0202M 2×10^6 U/ml).

RNA was removed by incubating samples with 15 µl of 2 µg/µl RNase A (Roche) for 45 min at 37 °C. Cross-link was reversed by incubating aliquots with 10 mg/ml proteinase K (Roche) at 65 °C overnight. DNA was then purified by AmpureXP SPRI beads 1× (Beckman Coulter; A63881) with two elution rounds of 100 µl each.

3way-4C. 3way-4C was performed using the UMI-4C protocol²⁹, with a modified size selection approach intended to maximize the number of products involving more than two ligation partners. Briefly, 5–10 µg of 3C template was sonicated at 4 °C using Bioruptor (Diagenode) for 5 cycles of 25 s on, 60 s off. The sonicated material was inspected on 10% agarose gel to assure size distribution of 700–900 bp. The sonicated DNA was then subjected to end-repair reaction: 20 µl 10× end-repair buffer and 10 µl end-repair mix (NEB E6050L) were added to the mix and incubated at 20 °C for 30 min. The DNA was cleaned up with SPRI (2.2×) beads and eluted in 76 µl EB (10 mM Tris–HCl pH 8.5). A-tailing reaction was performed with 10 µl NEB buffer 2, 4 µl Klenow fragment -3′-5′ exo (NEB M0212M), and 10 µl dATP 10 nM. Following A-tailing, the 5′ ends of the DNA were dephosphorylated with 2 µl of calf intestinal alkaline phosphatase (NEB M0290S) at 50 °C for 60 min. DNA was cleaned up with 2.0× SPRI beads and eluted in 67 µl EB. Template DNA aliquots of 1,000 ng were ligated with 5 µl Illumina indexed adapters 15 µM (final concentration 0.4 µM) using 10 µl quick ligase mix and 80 µl 10× buffer (NEB M2200). To release the non-ligated strand of the adaptor, the DNA was denatured at 95 °C for 3 min, immediately transferred to ice and cleaned up with 1.0× SPRI beads.

Two nested PCRs were used for UMI-4C library construction. The PCRs were performed in a final volume of 50 µl with the following reagents: 10 µl GoTaq Flexi buffer (Promega M792A), 3 µl MgCl 25 mM (Promega A3511), 25 µM dNTPs, 2 µl from a 100 µM primer pool targeting the regions of interest (final concentration of each bait varies depending on the pool size), 2 µl Illumina enrichment primer II 10 mM (0.4 mM final concentration), 1 µl GoTaq HotStart Polymerase (Promega M5005) and 200 ng UMI-4C DNA template. PCR program: one cycle of 2 min at 95 °C, 20 cycles of 30 s at 95 °C, 30 s at 56 °C, and 60 s at 72 °C, plus a final extension cycle of 5 min at 72 °C. Following the first PCR, products were cleaned up with

1.0× SPRI beads. The second PCR was identical to the first but contained a different set of primers that target a nested sequence between the 4C bait and the next downstream restriction site. The second primer is also attached with the Illumina I adaptor sequence. The program for the second PCR was the same as for the first but only went through 16–18 cycles.

The amplified DNA was cleaned up with 0.7× SPRI beads and the size distribution was inspected using TapeStation D1000 (Agilent Technologies). Typically, libraries had a size distribution around 800–1,000 bp. The libraries were then pooled to 10 nM and were multiplexed with other libraries for paired-end sequencing on the Illumina Miseq platform (300 cycles V2; 2×150).

C-walks. High molecular weight 3C template was isolated by gel excision of the upper band above the 10 kb marker using QIAEX II gel extraction kit (QIAGEN 20021). Serial dilutions (1:10) of size selected 3C template were performed to reach a final concentration of 10 pg/µl as the stock template. The REPLI-g single cell kit protocol for purified genomic DNA (QIAGEN 150343) was used following the instructions from the manufacturers with a few modifications. In short, all reagents were mixed at the same volumes and proportions specified by the manufacturers but splitting the final volume of a single reaction into ten individual ones, reducing the starting material amount to ~1 pg. To prevent contamination, all operations preceding the amplification step were performed under strict hygienic conditions (TC hood) and thorough DNA decontamination with DNAzap (ThermoFisher) and ultraviolet irradiation of laboratory utensils. A single C-walk amplification reaction consisted of ~1 pg of 3C material. Typically for 100 reactions, 10 µl of 10 pg/µl stock template were mixed into 16.56 µl of sterile TE buffer and denatured at room temperature for 3 min by mixing with buffer D1 (5.83 µl DLB buffer plus 20.83 µl scH₂O; both from REPLI-g sc kit). Denaturation of DNA was stopped by addition of 50 µl buffer N1 (7.5 µl Stop solution plus 42.5 µl scH₂O; both REPLI-g sc kit) and stored on ice. A final Master Mix was prepared by mixing 400 µl of reaction buffer (290 µl reaction buffer plus 90 µl scH₂O plus 20 µl Φ29 DNA polymerase; from REPLI-g sc kit). An aliquot of 5 µl of the master mix was taken as a negative control. The master mix and DNA template were then mixed and kept on ice. Aliquots of 5 µl from the master mix were isolated into single wells from 96-well plates, sealed and incubated for 2 h and 50 min at 30 °C, followed by a heat inactivation step of 3 min at 65 °C. Additional 45 µl of H₂O were added to each well after spinning down for a few seconds. The amplified material was cleaned up with 2.5× SPRI beads and resuspended in 100 µl.

C-walk library construction. Φ29-amplified material (300–700 ng per library) was sonicated in Bioruptor (program: 11 cycles of 5 s on, 90 s off) in 0.1 ml Bioruptor Microtubes (C30010015) to reach a DNA fragment size between 900–1,500 bp, peaking at 1,200 bp. The sonicated DNA (42.5 µl) was then subjected to end-repair reaction (5 µl 10× end-repair buffer plus 2.5 µl end-repair mix (NEB E6050L)) at 20 °C for 30 min. The DNA was cleaned up with 2.2× SPRI beads and eluted in 38 µl EB. A-tailing reaction was performed with 5 µl NEB buffer 2, 4 µl Klenow fragment -3′-5′ exo (NEB M0212M) and 10 µl dATP 10 nM for 30 min at 37 °C, cleaned up with 2.2× SPRI beads and eluted in 30 µl EB. DNA template was ligated with 5 µl Illumina indexed adapters 0.75 µM with 5 µl T4 quick ligase (NEB M2200) and 40 µl 10× ligation buffer for 15 min at 25 °C. Ligated template was then cleaned up with 1.3× SPRI beads and eluted in 41 µl.

Each indexed template was PCR amplified on a final volume of 50 µl with the following reagents: 1 µl dNTPs (10 mM), 2 µl Illumina enrichment primer I and II mix 10 mM, 1 µl Pfu ultra II Fusion HS DNA polymerase (Promega 600670), 5 µl Pfu ultra II Fusion buffer (Promega 600670) and 200 ng of Φ29-amplified material. PCR program: 2 min at 95 °C, 12 cycles of 30 s at 95 °C, 30 s at 55 °C, and 60 s at 72 °C and final extension cycle of 10 min at 72 °C. Following the PCR, products were cleaned up with 1.0× SPRI beads and DNA concentration determined by fluorometric quantitation using Qubit (ThermoScientific).

Libraries were pooled in groups of eight in equal quantities and ran on a 1% agarose gel, not loading more than 1,000 ng per lane. Gel was excised selecting for mean library size depending on the sequencing platform used. For shorter runs (HiSeq) libraries were selected 400–600 bp and a range of 800–1,200 bp for long runs (MiSeq). DNA was purified using minelute gel extraction kit (QIAGEN). Samples were pooled together and paired-end sequenced using the Illumina platform.

PacBio sequencing. Raw 3C material was cleaned up with 0.5× SPRI beads and standard PacBio libraries were prepared for K562 and mES cells. Specifically, K562 libraries, we observed a 3.12 kb sub-read N50 with a total of 1.5 Gb sequenced across 612 kb sub-reads. For the mES cells we observed a sub-read length N50 of 1.81 kb with 1.4 Gb of total sequencing over 882 kb sub-reads. Circular consensus sequence (CCS) for K562 showed an average read length of 1,769 bp; maximum 7,881 bp and 1,382 bp for mES cells; maximum 7,327 bp. Downstream analysis of PacBio reads made use of the same pipeline as shotgun derived sequences

(see below) with the addition of an extra splitting step where reads larger than 160 bp were split again into 80 bp sub-reads. Once short reads were mapped, sub-reads were merged based on their mapping coordinates. Chains of mapping points were directly assembled into C-walks as reads derived from PacBio do not need to go through the reassembly step inherent of a shotgun approach.

Additional imported data sets. We imported HiC and one-dimensional epigenomics data for K562, and one-dimensional epigenomics data for mES cells as listed in Supplementary Table 3.

Analysis code. Code recreating all reported analysis, as well as processed data files, is available from our website: http://compgenomics.weizmann.ac.il/tanay/?page_id=99.

Hi-C mapping and filtering. We mapped paired-end sequences to DpnII fragment ends as previously described¹⁰, with small modifications. Reads were first segmented into postulated DpnII (sub-fragments) whenever a GATC subsequence was observed. Bowtie2 was then used to map segments to the respective genome assembly (mm9 or hg19). Uniquely mapped sequences with sufficient quality (MAPQ > 30) were first tested for complementarity between read 1 and read 2 suffixes, and merged if such overlap exists. Merged mapped coordinates were then projected onto a precomputed set of fragment-ends using the mapped coordinate and strand, and subsequent pairs were used to build the Hi-C matrix. Pairs of fragment ends that are separated by less than 1,000 bp were considered as non-digested products and discarded from further processing. For mES Hi-C we acquired a total of 145 million pairs. For K562 Hi-C we reprocessed published data¹¹ including 405 million pairs.

Defining TAD structure and annotating active/inactive TADs. To define TAD structure from Hi-C we use an approach similar to the one previously described¹⁰. Briefly, given a Hi-C matrix, we define the insulation score (for a scale of s bp) at a coordinate X by computing:

$$X_{\text{triangle}}(x,s) = \text{number of contacts between any pair of elements in the interval } (x-s, x+s) \\ X_{\text{insul}}(x,s) = \log_2\{(X_{\text{triangle}}(x,s) - X_{\text{triangle}}(x+s/2, s/2) - X_{\text{triangle}}(x-s/2, s/2))/X_{\text{triangle}}(x,s)\}$$

We use a threshold of $X_{\text{insul}} < T$ ($T = -2.769$ K562; $T = -2.645$ mES) to create a grid of insulation points, and identify TADs as elements between such insulation points. We filter TADs smaller than 250,000 bp.

To classify TADs into active and inactive clusters, we computed the mean time of replication, mean H3K4me3 level at the TSS and additional one-dimensional epigenomic features for each TAD. TADs with early replicating statistics were classified as active and TADs with later replicating statistics as inactive. This approximation is consistent with other metrics (for example, gene expression), and with inference of an A/B compartment directly from the map (data not shown).

Probabilistic normalization and visualization of pairwise Hi-C maps. To normalize and visualize Hi-C data at high resolution we applied an approach that computes an enrichment score for each observed Hi-C contact as follows.

First, to visualize a chromosomal region (minimum (min), maximum (max)) we first identify all contacts $M = (x_i, y_i)$ within a large intra-chromosomal submatrix such that $\min' < x_i, y_i < \max'$ on a range (\min', \max') where $\min' = \min - (\max - \min) \times 2$ and $\max' = \max + (\max - \min) \times 2$. Let N be the total number of contacts we extract. We note that we use M as a symmetric matrix (duplicating contacts (X, Y) to (Y, X)).

Second, we identify the set of coordinates from the participating fragment ends in the set M to define $U = \{u_i\}$. We compute the number of contacts per element in the matrix, denoted $n(u_i)$.

Third, we estimate the contact decay regime by binning pairs $(x_i - y_i)$ according to their logarithmic distance bin:

$$b_i = \text{floor}(\log_2(100 + |x_i - y_i|))$$

and computing:

$$r(d) = \#\{i \text{ s.t. } b_i = d\} / N$$

We then correct this estimation to account for edge effects (that is, only the pairs at the corners of the matrix can have maximal distances):

$$r'(d) = 1/Z \times r(d) / \{\#(j, j') \text{ with } \text{floor}(\log_2(100 + |u_j - u_{j'}|)) = d\}$$

Here Z is a normalization coefficient and counting is going over all pairs of coordinates in U .

Fourth, we generate a randomized set of contacts M_r using a Markov chain Monte Carlo approach that preserves the total number of contacts per element as well as the overall decay trend r' . This is done through a metropolis algorithm as

follows: (1) initialize $M_r = M$; (2) repeatedly sample two contacts $(x_i, y_i), (x_{i'}, y_{i'})$; (3) compute the probability of the sampled contacts: $P_1 = r'(d_i) \times r'(d_{i'})$ (d_i and $d_{i'}$ are the logarithmic distance bins for the contacts); (4) compute the probability of the shuffled contacts $(x_i, y_{i'}), (x_{i'}, y_i)$ $P_2 = r'(d'_{i'}) \times r'(d'_{i'})$ (again, the d' values are the log-distance bins of the shuffled contacts); (5) accept the shuffle with probability $\min(1, P_2/P_1)$ and, when accepting, update M_r by removing the original contact and adding the shuffled one; (6) continue sampling until $N \times K_{\text{burn}}$ shuffles are accepted; and (7) symmetrize the final M_r by generating the symmetric contact Y, X for each contact X, Y . We therefore end up with a symmetric matrix having $2N$ contacts.

Fifth, we compute for each contact in the original contact matrix M , the Euclidean distances (in two-dimensional space) to the nearest K_{near} neighbours, generating a series $D^i = (d_1, \dots, d_{K_{\text{near}}})$. We also compute the distances from this contact to the $K' = 2 \times K_{\text{near}}$ nearest neighbours in the pooled randomized matrix $MR = D'_r = (d'_1, \dots, d'_{K'})$. Note that the original contact is unlikely to be present in M_r . The (pairwise) enrichment score $s(i)$ is defined as the Kolmogorov-Smirnov D statistics derived by comparing the distributions of D^i and D'_r .

Last, to avoid additional edge effects, we plot and use data for the original (min,max) range, and discard scores that were computed for the larger range as defined above.

For the Hi-C analysis shown in this paper, we used $K_{\text{burn}} = 10$, $K_{\text{near}} = 100$.

Mapping, filtering and assembly of C-walks. C-walk libraries contain paired-end sequences with a well-specific index, which are processed by the following pipeline steps.

First, de-multiplexing of well barcodes (typically as part of the de-multiplexing pipeline of the sequencer).

Second, mapping read-pairs of each label using the Hi-C mapping pipeline discussed above. Note that by splitting reads over DpnII sequences, we can detect more than one contact in one read-pair. Following this step we generate a list of fragment-end pairs per well.

Third, aggregating the data to generate a list of distinct fragment-end pairs and the number of reads supporting these, separately for each well.

Fourth, eliminate potential cross-well contamination: for each fragment end-pair present in the entire amplification batch, find the well for which its read coverage is maximal, and define this coverage as r . Eliminate all other appearances of the pair if their coverage is less than $0.1 \times r$ (typically we observe no overlap of pairs between wells).

Fifth, build the C-walk graph $G = (V, E)$ for each well: nodes in this graphs are fragment ends, and arcs are added for each observed contact pair. Label arcs using read coverage $r(E)$.

Sixth, resolve the C-walk graph. This is a simple algorithm that relies heavily on the idea that almost all fragments in the well are present in at most two real ligations, and that these ligations should be supported by a much larger number of read pairs than spurious ligations that were introduced by amplification or sequencing errors. The first step is to initialize an empty graph $P = (V, E')$ on nodes V , with empty arc set E' . The initial degrees in the graph are set to $\deg(V) = 0$. The second step is to iterate over all arcs (x, y) in E , ordered by read coverage $r(E)$. The third step is that if the $\deg(x) < 2$ and $\deg(y) < 2$ (that is, both X and Y are not yet connected to two nodes), add (x, y) to E' and update the degrees: $\deg(x) = \deg(x) + 1$, $\deg(y) = \deg(y) + 1$. Otherwise, (x, y) is discarded. The graph P consists of only paths and (very rarely) cycles. Report each connectivity component in the graph as a C-walk.

Last, within C-walks, merge hops that connect fragment ends that are less than 1 kb apart (since these are likely representing incomplete DpnII site digestion).

While this algorithm is very simple and may compute suboptimal C-walks in cases where there is substantial cross-C-walk noise, the approach it implements is conservative, and increases the probability that inferred C-walk hops are true positives. Missed hops, and occasional assembly errors imposed by the amplification process are shown to be of limited impact in our analysis of linear DNA reassembly (Extended Data Fig. 1). Additional controls with mixed human-mouse DNA 3C templates further support the absence of cross C-walk noise (Extended Data Fig. 1f).

C-walk simulations. To simulate C-walks from the pairwise Hi-C distribution, we constructed a graph based on the restriction fragment space and arcs connecting them based on empirical Hi-C data. For computational efficiency, we randomly sampled 5 contacts for each fragment end. We then generated random walks in the Hi-C graph, maintaining the distribution of the number of hops per walks as estimated from the C-walks data. This resulted in a synthetic data set that we analysed using the same approach used for real data, as shown in Extended Data Figs 2f, g, 3h and 4i, j.

3way-4C mapping and filtering. 3way-4C read pairs are similar to standard Hi-C/4C products, mapping to one, two, three, or more genomic loci and reporting

on proximity events linking them. We processed read-pairs from 3way-4C experiments using the following procedure.

De-multiplexing. Using read 1 prefix, we grouped read pairs according to their respective viewpoint primers.

Filtering non-specific amplifications. Using a designed pad sequence separating the viewpoint primer and the adjacent DpnII restriction site, we remove products that represent non-specific amplifications.

DpnII segmentation. We identified GATC sequences in read 1 and read 2, and generated a set of sequences representing (parts of) restriction fragments.

Mapping. Using bowtie 2 and standard parameters, relying on the above segmentation to avoid chimaeric sequences. We retained for further processing only uniquely mapped products (mapping score >30). We note that although the bait fragment is present and uniquely mapped for all de-multiplexed reads by design (as it contains the primer sequence), only the requirement for an additional pad sequence following the primer ensures the viewpoint fragment was indeed part of the ligated sequences.

Triplet identification. We grouped all mapped segments that originated from the same read pair and merged overlapping segments from read 1 and read 2. We then identified all read pairs that define at least three fragments (the first three if more than three are identified), and generated a list of triplets (bait, fend1, fend2).

Eliminating PCR duplicates using a UMI strategy. Each 4C read pair is marked by the sequence prefix of read 2 that is determined by a (random) sonication breakpoint. We filtered PCR duplicates by analysing all read pairs contributing to the same triplet of ligation partners (bait, fend1, fend2), and collapsing together all cases that have a common sonication point (within one bp) at fend2.

Merging data from multiple amplifications. We typically perform several PCR amplifications for each experiment, which are de-multiplexed and processed separately. After the triplet identification step we can merge data on the triplet molecules from each PCR index to generate the triplet data set for further statistical analysis as discussed below.

Probabilistic normalization and visualization of 3way-4C triplets. Given a list of triplets for a view point (V, X_i, Y_i) we characterize synergistic tri-way contacts by sampling randomized triplets as described below and comparing the observed triplet densities to the randomized data.

Bait triplets background model. We assume triplets are generated from two pairwise contacts independently. Since we model contacts anchored on the bait, once the pairwise contact associates V with a locus X , we assume that this indicates the chromosomal neighbourhoods of X and V are now linked. We therefore sample the second pairwise contacts from either the pairwise contacts of X or those of V using a fixed mixing parameter p . The mixing parameter is estimated from C-walk data by extracting pairs of consecutive hops $(X1, X2, X3)$ such that $|X1 - X2| > 50$ kb, and counting cases such that $|X2 - X3| < 5$ kb and cases such that $|X1 - X3| < 5$ kb, setting p according to the ratio between them. This resulted in preferring sampling from X neighbourhood over the neighbourhood of V by a ratio of 2:1.

Sampling from V and X chromosomal neighbourhoods. Sampling contacts from the neighbourhood of V is straightforward since the data (V, X, Y) provide us with a high complexity 4C contact profile for V . To sample contacts from the X pairwise neighbourhood, we use a Hi-C matrix and identify 200 contacts (U_j, W_j) with minimal $|X - U_j|$, and sample contacts from $(X, X + (W_j - U_j))$. This strategy provides

adequate correction for the pairwise contact preferences of any genomic element that was contacting the bait V , provided sufficient Hi-C coverage is available to ensure the approximation neighbours of X is accurate.

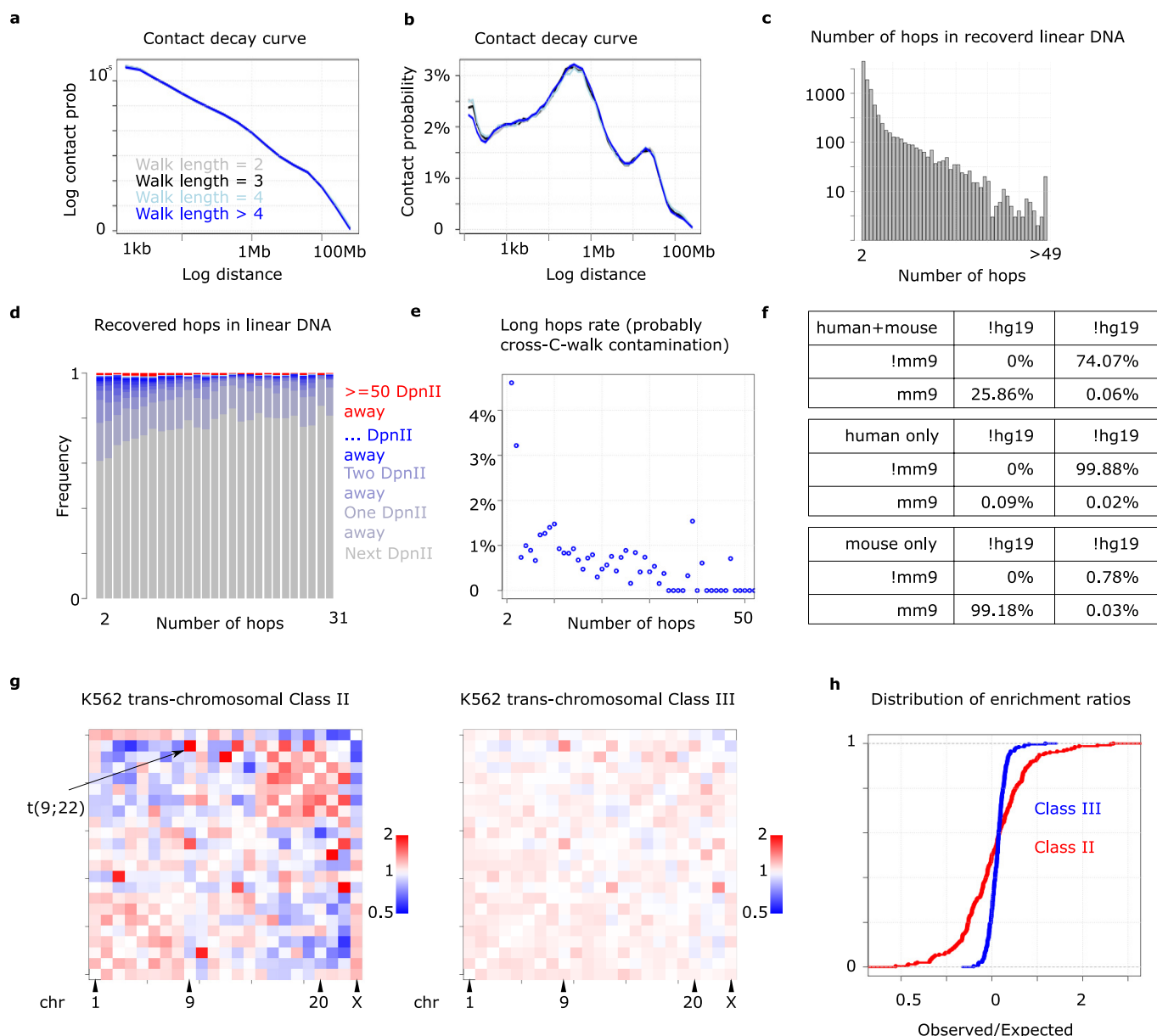
Clipping data. We perform all analysis below by clipping 3way-4C and Hi-C data to a region around the bait (2 Mb on each side). We note it is important to ensure the same clipping policy is used for both the triplet data and the Hi-C data, such that the nearest neighbours (U, W) sample a similar distance range as represented in the triplet clipped sub-matrix. We also eliminated all contacts with $|X - Y| < 3$ kb, since these showed strong additional biases that are not informing on true long range hubs. Such biases will have to be further characterized with analysis of multiple baits and conditions.

Generating the randomized triplet data set. We use the ideas outlined in the first three steps as follows. Given (V, X_i, Y_i) , we sample $(1 - p) \times N_{\text{rand}}$ permutations (perm) and add them to the randomized triplets set $(V, X_i, Y_{\text{perm}(i)})$. Then we go over all (V, X_i) pairs, and for each we sample $p \times N_{\text{rand}}$ indices $j(k)$ uniformly from 1–200 with replacement. We then add contacts $(V, X_i, X_i + (W_{j1} - U_{j1})), \dots, (V, X_i, X_i + (W_{jN_{\text{rand}}} - U_{jN_{\text{rand}}}))$, where (U, W) are defined as the 200 Hi-C contacts closest to X_i as described above. Finally we symmetrize the resulted matrix (X, Y) and use it as the randomized data.

Computing hub-scores. We perform K -nearest neighbour search for each element in the triplet data set. In this case, we use $K_{\text{near}} = 100$ for T , and $K'_{\text{near}} = 100 \times 2 N_{\text{rand}}$ for the randomized triplets set in K562 data, and $K_{\text{near}} = 60$ for mES baits (since these had lower triplet counts overall). Recall that the K nearest neighbours (KNN) strategy is comparing the distributions of two-dimensional Euclidean distances between the observed triplet (X, Y) and its nearest neighbours in the observed and randomized data, computing Kolmogorov–Smirnov D statistics to visualize positive (higher density in observed data) and negative (lower density in observed data) enrichments.

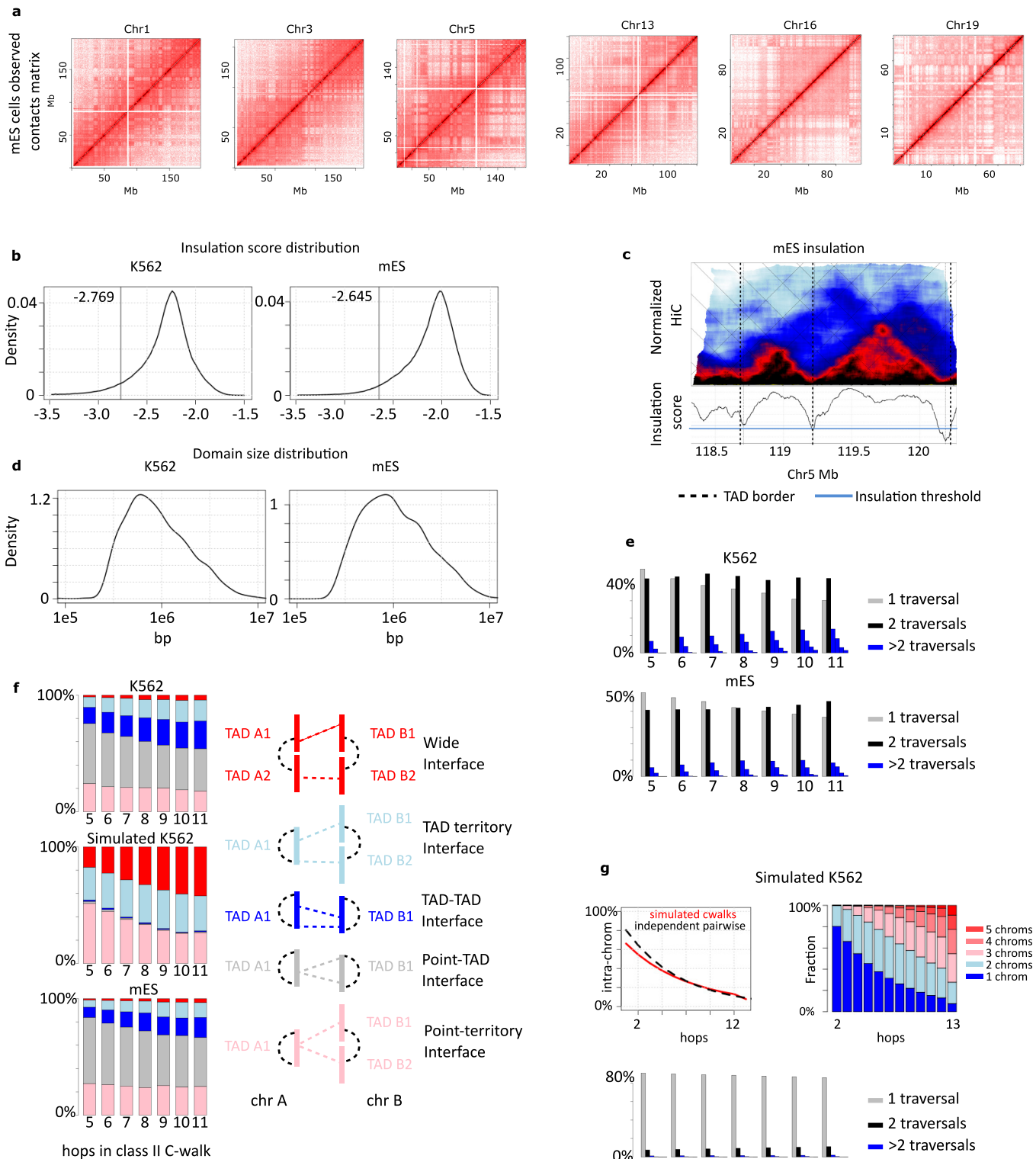
Analysing score distribution and robustness. We validate our scoring scheme by testing score distributions on data randomized by sampling from the model described above. This validation therefore accounts for (only) the technical correctness of the procedure, and also quantifies the effect of multiple testing (which is difficult to describe analytically given the high degree of dependency between tests). We set the scoring colour scale in the triplet figures such that the false discovery rate of the scores, as estimated from such synthetic data, is controlled (see Extended Data Fig. 5c).

28. Geula, S. *et al.* m⁶A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347**, 1002–1006 (2015).
29. Schwartzman, O. *et al.* UMI-4C for quantitative and targeted chromosomal contact profiling. *Nature Methods* **13**, 685–691 (2016).
30. Ryba, T., Battaglia, D., Pope, B. D., Hiratani, I. & Gilbert, D. M. Genome-scale analysis of replication timing: from bench to bioinformatics. *Nature Protocols* **6**, 870–895 (2011).
31. Hansen, R. S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl Acad. Sci. USA* **107**, 139–144 (2010).
32. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the β -globin locus. *Genes Dev.* **20**, 2349–2354 (2006).
33. Deng, W. *et al.* Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* **158**, 849–860 (2014).



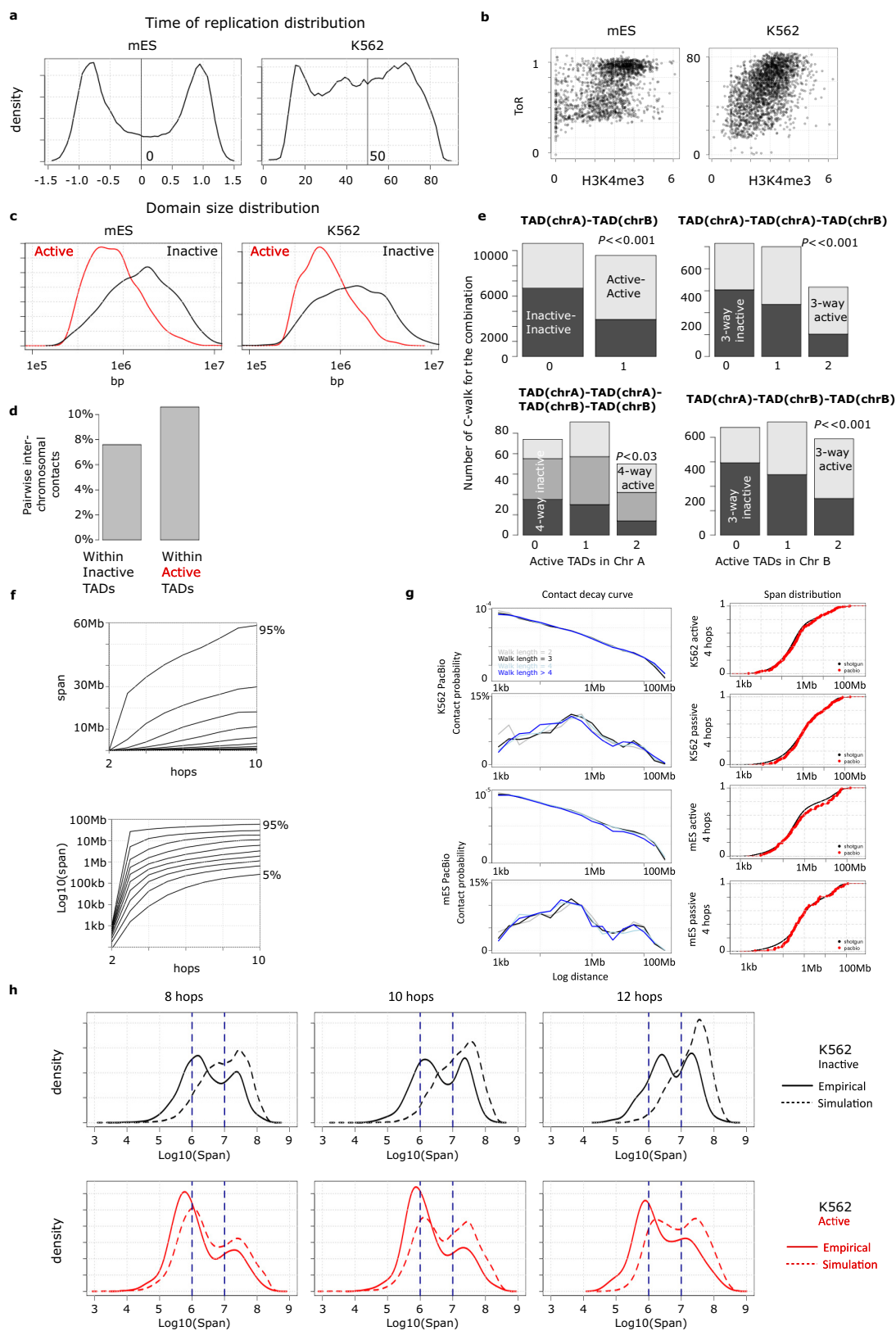
Extended Data Figure 1 | C-walk profiling. **a**, Log probability of contact as a function of log linear genomic distance, shown based on data from C-walks of length 2,3,4 and longer than 4. When quantifying pairwise contact probability from a C-walk, all hops are considered. **b**, Similar to **a**, but showing contact probability for logarithmic bins (without dividing by bin size). Overall these data show that contacts occur with probabilities that have a similar scale for close, medium and far ranges. However, the intra-domain distance range contacts (<1 Mb) are enriched 2–3 fold above the 10 Mb distance range. **c**, Distribution of the number of hops derived when analysing linear genomic DNA using the full C-walk protocol. **d**, Bar graph depicts the probability of C-walks (of increasing length) derived from linear genomic DNA to jump over 0, 1 and up to 20 and above DpnII restriction fragments. We note that some jumps occur due to short fragment sizes, and other due to imperfect amplification and sequencing. **e**, The probability of inferring a long hop (over 50 fragments) in linear genomic DNA C-walks. This provides estimation on the total level of noise (cross C-walk contamination) expected when performing C-walk analysis on 3C DNA. **f**, Cross-species controls. Human- and mouse-derived 3C material was mixed and processed together to measure the level of cross-C-walk contamination using 3C template instead of

genomic DNA. Reads were mapped to a hybrid genome comprising the hg19 assembly appended by the mm9 assembly. The levels of reads that mapped to both genomes is minute and comparable to the levels found when mapping human-only and mouse-only C-walks to the hybrid assembly confirming the fact that C-walk cross contamination is not a major source of noise in the shotgun approach. **g**, We used C-walks from class II with four or more hops (left) and C-walks from class III (right), to extract all inter-chromosomal hops and reconstruct a chromosome-to-chromosome contact matrix. We compute the expected number of pairwise chromosomal contacts for each pair by multiplying the total number of inter-chromosomal contacts for the two chromosomes and scaling to the sum of observed contacts. Shown are colour-coded enrichment ($\log(\text{observed}/\text{expected})$) values for each chromosome pair, indicating, for example, a clear t(9;22) signature (as expected in K562) in class II data. Class III C-walks recreate a nearly random contact preference among all chromosomes. The residual signal in the class III matrix suggests a low rate of non-spurious hops that are misclassified as class III C-walks. **h**, The cumulative distributions of the $\log(\text{enrichment})$ ratios for both classes is also shown.



Extended Data Figure 2 | Hi-C and TAD inference. **a**, We performed 4-cutter Hi-C in mES cells, generating 145 million distinct contacts. Shown are large-scale Hi-C maps derived from this data. **b**, Genome-wide distribution of insulation intensities in mES and K562 cells, using a 200 kb scale (Methods; computed as described previously¹⁰). The threshold values we used in order to demarcate TADs are marked as vertical lines. **c**, Overlay of the analogue insulation intensity level on a region of chromosome 5 in mES cells, illustrating TAD calling. Our analysis suggests that there is no single canonical definition of TADs, since loops occur at several scales, but the threshold we select is providing a good reference scale that demarcate TADs with typical sizes of 1–2 Mb. **d**, Distributions of TAD sizes. **e**, Distribution of the number of inter-chromosomal traversals for class II C-walks. **f**, We classified all class II C-walks with two or more

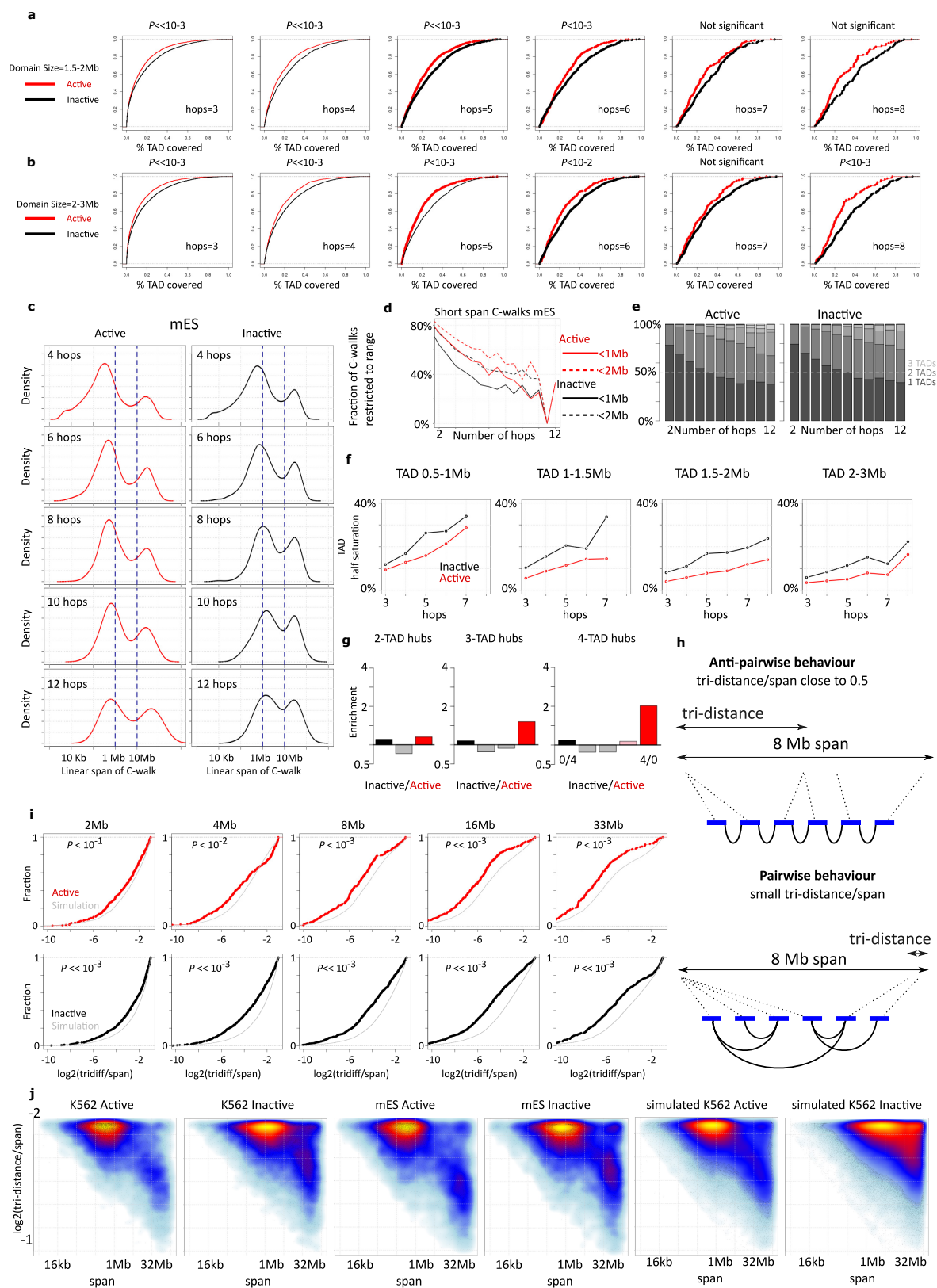
inter-chromosomal-traversing contacts into five subclasses determined by the properties of the two contacting elements on both chromosomes ('point', both contacts are contained within a 1 kb range; 'TAD', restricted to one TAD; 'territory', spanning more than one TAD). We found that only a small fraction (<3%) of the interfaces coupled the two chromosomes through contacts between multiple TADs at both chromosomes. On the other hand, most interfaces either linked the chromosomes through two contacting TADs (grey and blue), or associated one TAD in one chromosome with elements from more than one TAD in the other chromosome. Data from simulated, pairwise independent C-walks are provided for control. **g**, Fraction of intra-chromosomal ligations, distribution of the number of chromosomes visited and the size of inter-chromosomal interface for simulated, pairwise independent C-walks.



Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | TAD inference statistics and C-walks derived by PacBio sequencing. **a**, Distributions of mean time of replication (ToR) across TADs. Note that normalization of ToR in mES³⁰ and K562 cells³¹ is different. **b**, Correspondence between ToR and mean enrichment of H3K4me3 on transcription start sites (TSSs). **c**, Distribution of TAD sizes, shown separately for the active and inactive classes. **d**, Probability of pairwise inter-chromosomal hops in C-walks (excluding class III), estimated separately from inactive (left) and active (right) source loci (regardless of the TAD type at the target locus). **e**, For each class II C-walk we counted the number and types (active/inactive) of TADs visited on each of the two chromosomes. The bar graphs summarize these statistics separately for walks visiting one TAD on each chromosome (top left), one TAD in one chromosome and two in the other (right), and two TADs on each chromosome (which are rare as suggested by the data in Extended

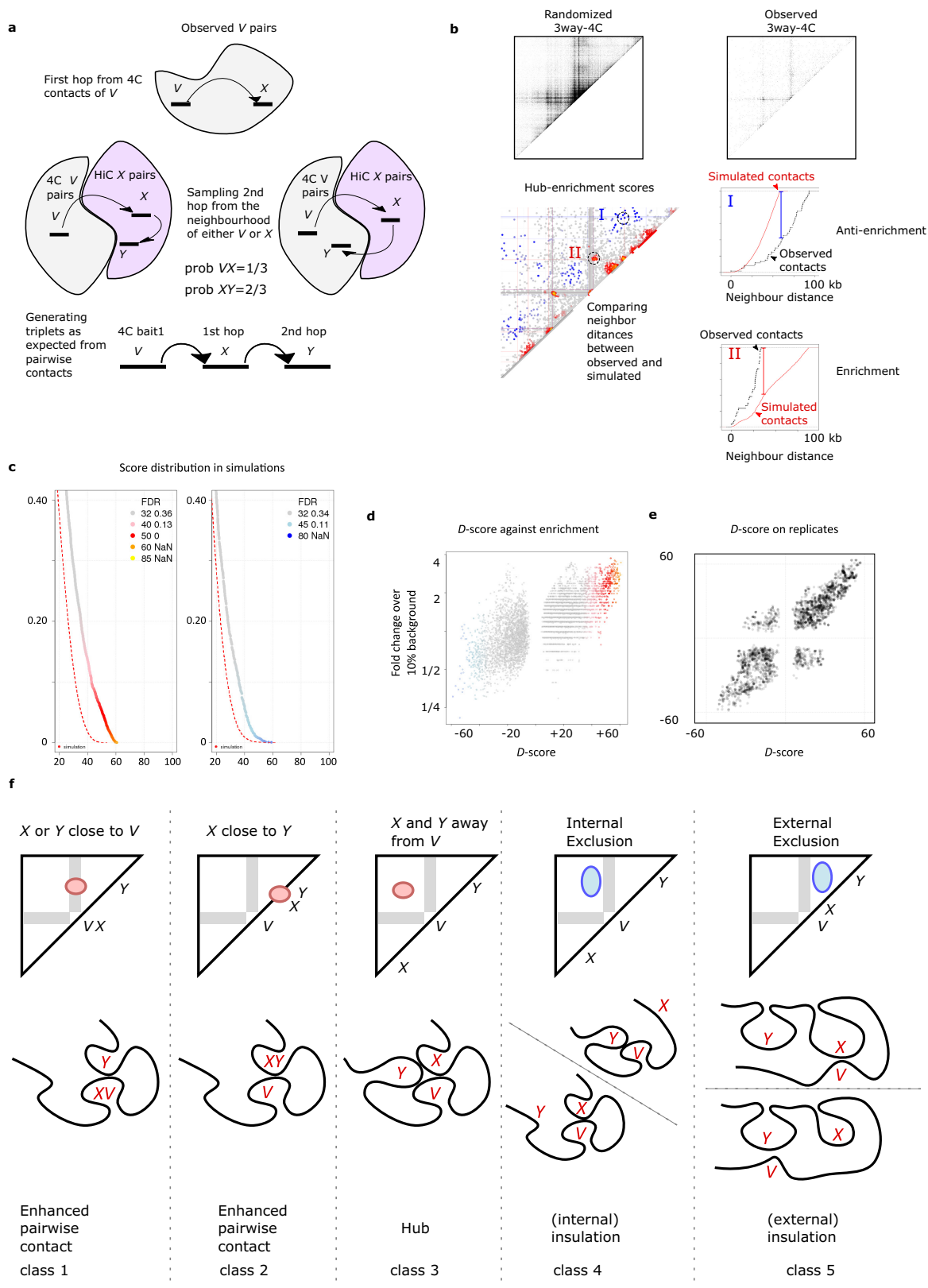
Data Fig. 2f). Each bar shows the breakdown of the number of active TADs visited in the second chromosome, conditioned on the number of active TADs visited in the first chromosome. *P* values are based on a chi-squared test, showing that composition of TAD types in the inter-chromosomal interfaces is non-random and prefers association of multiple active TADs or multiple inactive TADs. **f**, Shown are the 5–95% span quantiles (in steps of 10%) for C-walks with increasing number of hops (linear scale, top; log scale, bottom). **g**, Comparison of C-walks derived from our shotgun sequencing procedure (indirect) versus direct PacBio sequencing (direct). Statistics on the probability of pairwise contact given genomic distance (left), and the span of C-walks on four hops (right) corroborate the integrity of our C-walk assembly process. Note data are shown for both K562 and mES cells. **h**, Distributions of spans for observed and simulated C-walks in active (top) and inactive (bottom) TAD contexts.



Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | C-walk span supporting statistics and C-walk results on mES cells. **a, b**, Shown are cumulative distributions of the relative span (C-walk span divided by TAD size), for intra-TAD C-walks of increasing length (left to right), classified based on the type of TAD (active and inactive). Data are shown for TADs of two size ranges with sufficient statistics. *P* values are based on a Kolmogorov–Smirnov test. This supports Fig. 2g. **c–g**, Similar to Fig. 2d–g, j but using mES cells data. **h**, C-walks of more than three hops can be characterized as pairwise when they visit two chromosomal neighbourhoods but not any position in between them. Quantitatively, we define the tri-distance of a C-walk as the maximal distance of any hop from the minimum and maximum positions of the walk and analyse the distributions of the ratios between tri-distance

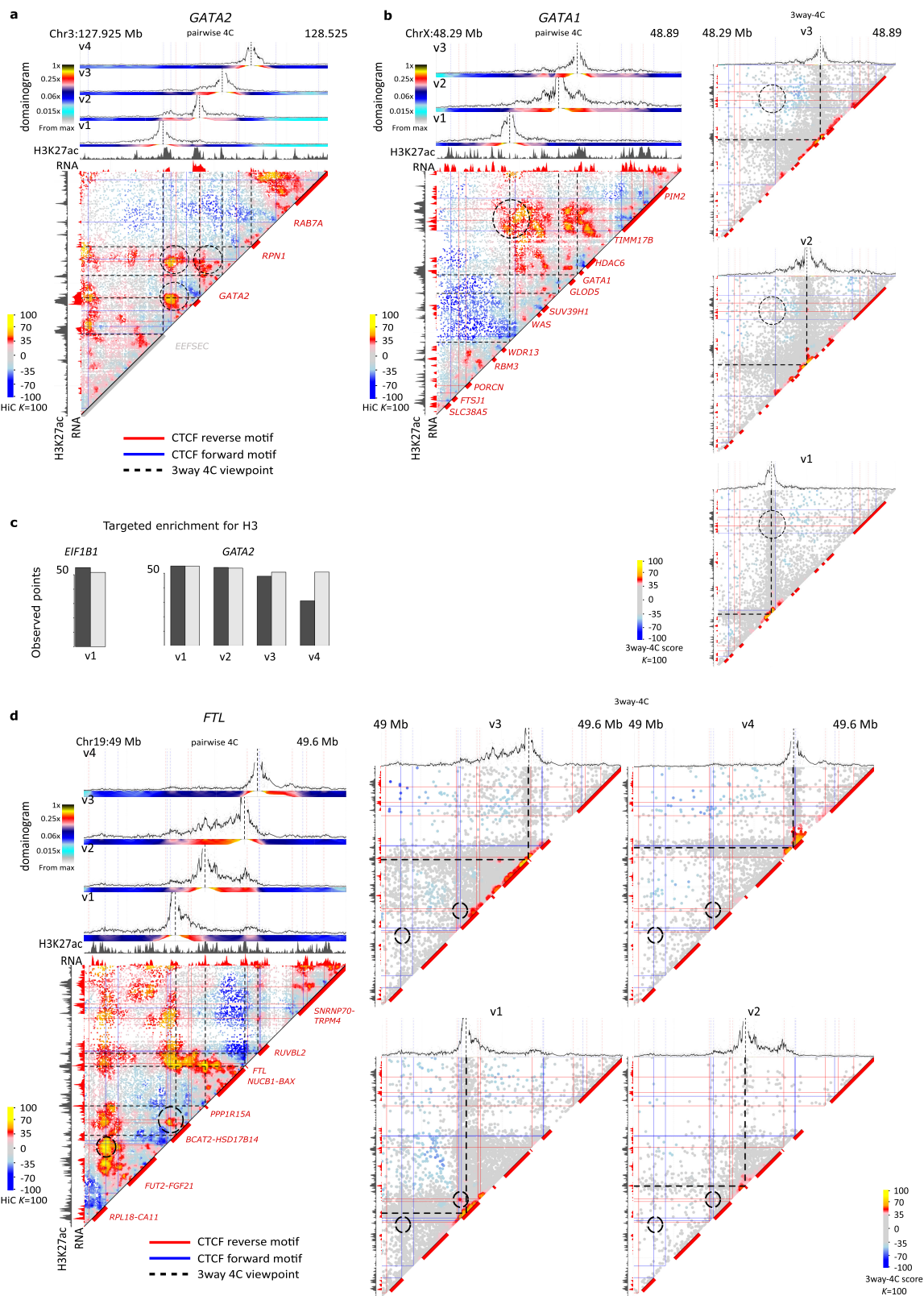
and span for C-walks of six hops and more. Low ratios indicate a pairwise interface, while ratios around 0.5 represent maximally non-pairwise structure. We observed that high-span C-walks are enriched for cases with tri-distance to span ratio of 0.1 and less. On the other hand, C-walks with shorter spans (for example, less than 1 Mb) usually do not show up as strongly pairwise, possibly since the sensitivity of 3C is limited for short-range contacts. **i**, Shown are cumulative distributions of observed and simulated tri-diff scores on C-walks stratified by span (from 2 Mb to 32 Mb). *P* values are based on Kolmogorov–Smirnov statistics. **j**, Heatmaps representing the distribution of tri-distance scores for C-walks stratified by span.



Extended Data Figure 5 | See next page for caption.

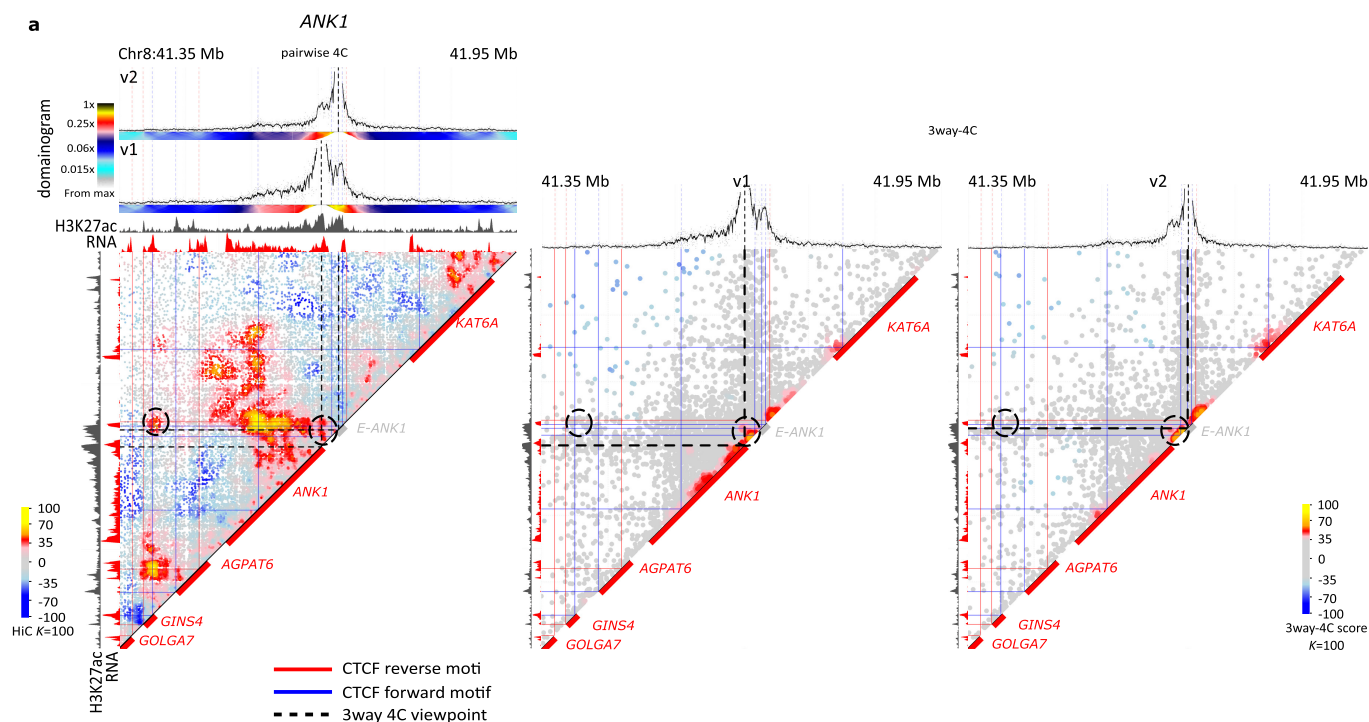
Extended Data Figure 5 | 3way-4C analysis. **a**, 3-Walk generation scheme. A first hop from the bait V to X is sampled from the pairwise chromosomal neighbourhood around V . This hop indicates that the neighbourhoods of V and X are now physically coupled. The second hop can therefore be sampled from either V (right) or X (left) pairwise neighbourhoods. The sampling ratio between the two neighbourhoods is determined from C-walk data. Note that if (V, X, Y) is a true synergistic hub, the probability of observing it should be higher than predicted by such independent sampling of pairs (V, X) , (V, Y) and (X, Y) . **b**, Analysis procedure. The input (right) is a set of 3-walks originating from the viewpoint bait V summarized in a matrix (X, Y) . Simulated (X, Y) triplets (left) are generated by sampling from the pairwise chromosomal neighbourhood as outlined in **a** (see Methods for complete details). To compare the observed and simulated triplet distributions, we compute for any observed triplet $(V-X-Y)$, the K nearest observed contacts (in Euclidean space), and K' (K scaled by the size of the sample) nearest simulated contacts. Comparison is done using Kolmogorov–Smirnov D statistics on the cumulative distributions of distances. If distances are significantly shorter in the observed data (distributions II), we annotate the triplet as synergistic (colour-coded red–yellow). If the distances are significantly shorter in the randomized data (distributions I), we annotate the triplet as insulated (colour-coded blue). **c**, Comparing score distribution on observed and simulated data. Shown are empirical cumulative density functions for positive (right) and negative (left) D statistics using data from the *EIF1B* locus (Fig. 3b). Colour coding is the same as the one used in the visualization of triplet matrices. **d**, We defined the background range for each point in the

EIF1B data set as the 0.1 quantile in the K' nearest neighbour within the simulated data set. We compute the fraction of contacts in the observed K nearest neighbour that are within this range, and derive an enrichment ratio by dividing the observed fraction by the background fraction (0.1). Shown is a comparison of the D score and the enrichment ratio (note the y axis is in log scale). We note that setting the background level at 10% is not always appropriate, so we use the D score (which is not pre-defining a background level) directly for visualization. **e**, Comparing D scores for replicate experiments in the *EIF1B* locus. **f**, Interpretation. Points with similar observed and simulated statistics (annotated as grey in the matrix), represent cases where pairwise 3C is predicting the behaviour of 3-walks. We note that this can be a consequence of limited 3C resolution. It can also suggest that within individual nuclei the contacts V, X and V, Y are indeed occurring independently. When enrichment is observed, we can classify it into at least three types. Cases in which enriched triplets are bringing together two loci that are either linked with each other (class 1), or linked on one end to the viewpoint (class 2), can be interpreted as a pairwise contact with specific additional local structure (for example, see Fig. 3c), or as restricting contact with additional loci given an existing pairwise contact. Note that in any case, the analysis excludes contacts that occur within less than 3 kb, so such local structure should be understood on a scale of few kilobases. Only when the hotspot involves three distinct loci can we hypothesize that a true long-range hub is observed (class 3). In addition, 3-walk distribution can also identify insulation between X and Y , which can be observed when interrogating from within the insulated loops (class 4) or outside of these (class 5).

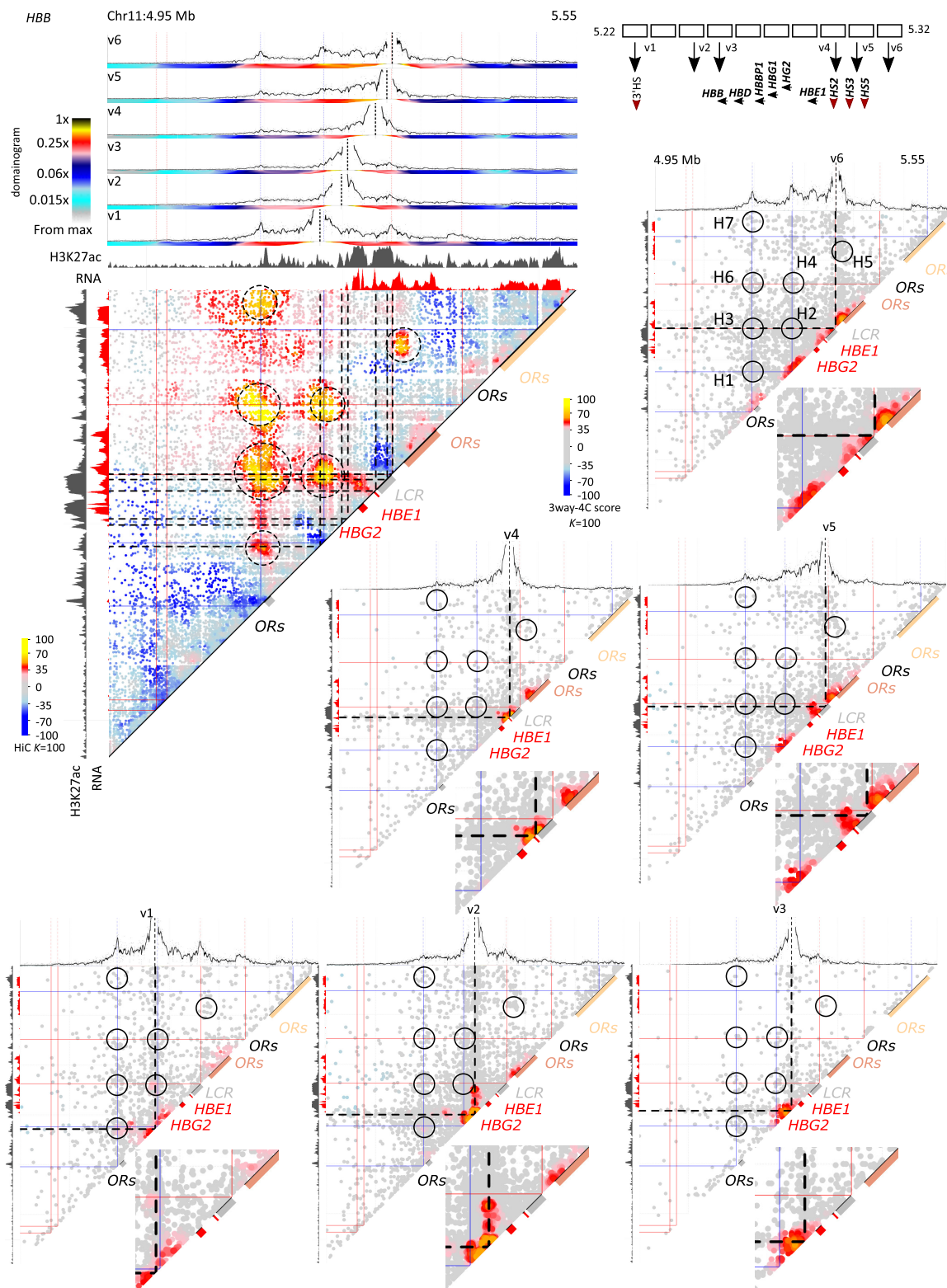


Extended Data Figure 6 | Hi-C and 3way-4C analyses for the *GATA2*, *GATA1* and *FTL* loci. a, Hi-C maps showing contact structure at the *GATA2* regions. The locus is flanked by two loops (H1 and H2), and these are shown by Hi-C to be frequently fused to form a bigger loop (H3). **b**, Hi-C and 3way-4C analysis at the *GATA1* locus. Note that this gene-dense region is qualitatively different from the *GATA2* or *EIF1B* loci, and is characterized by complex pairwise contact landscapes as shown by Hi-C. This suggests potential enrichment of many enhancer- or promoter-synergistic hubs. Nevertheless, the data suggest tri-way contacts are predicted by pairwise contact distributions around this region.

c, Bar graphs show observed and expected triplet contacts for H3 loops on *EIF1B1* and *GATA2* loci, supporting a lack of synergistic hub enrichment. **d**, Hi-C and 3way-4C analyses for the *FTL* locus. *FTL* is highly expressed in K562 cells, and is located in a region with many additional active genes. Multiple long-range contacts are observed in pairwise Hi-C analysis of this region. 3way-4C enrichment maps show that the pairwise models are generally consistent with the data, with some enrichment for contacts bringing together the *BCAT2-HSD17B14* locus with the encapsulating CTCF loop (v1).

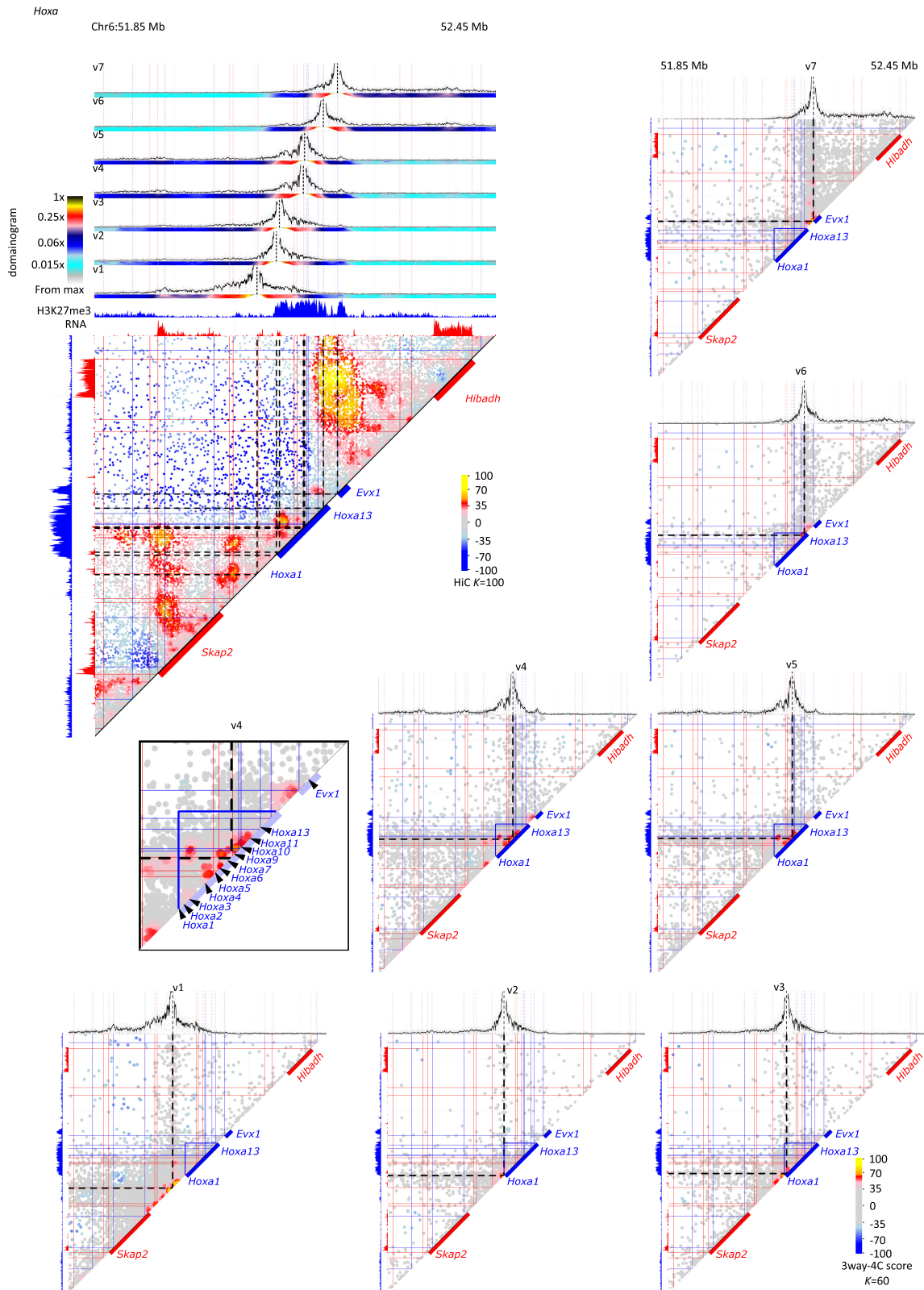


Extended Data Figure 7 | Hi-C and 3way-4C analyses for the *ANK1* locus. Similar to Fig. 3a, b. A complex contact structure around the *ANK1* locus is depicted by Hi-C analysis. 3way-4C on two baits suggests possible extended pairwise contact with the expressed *ANK1* and *KAT6A* loci, but offers no support for long-range synergistic tri-way hubs.

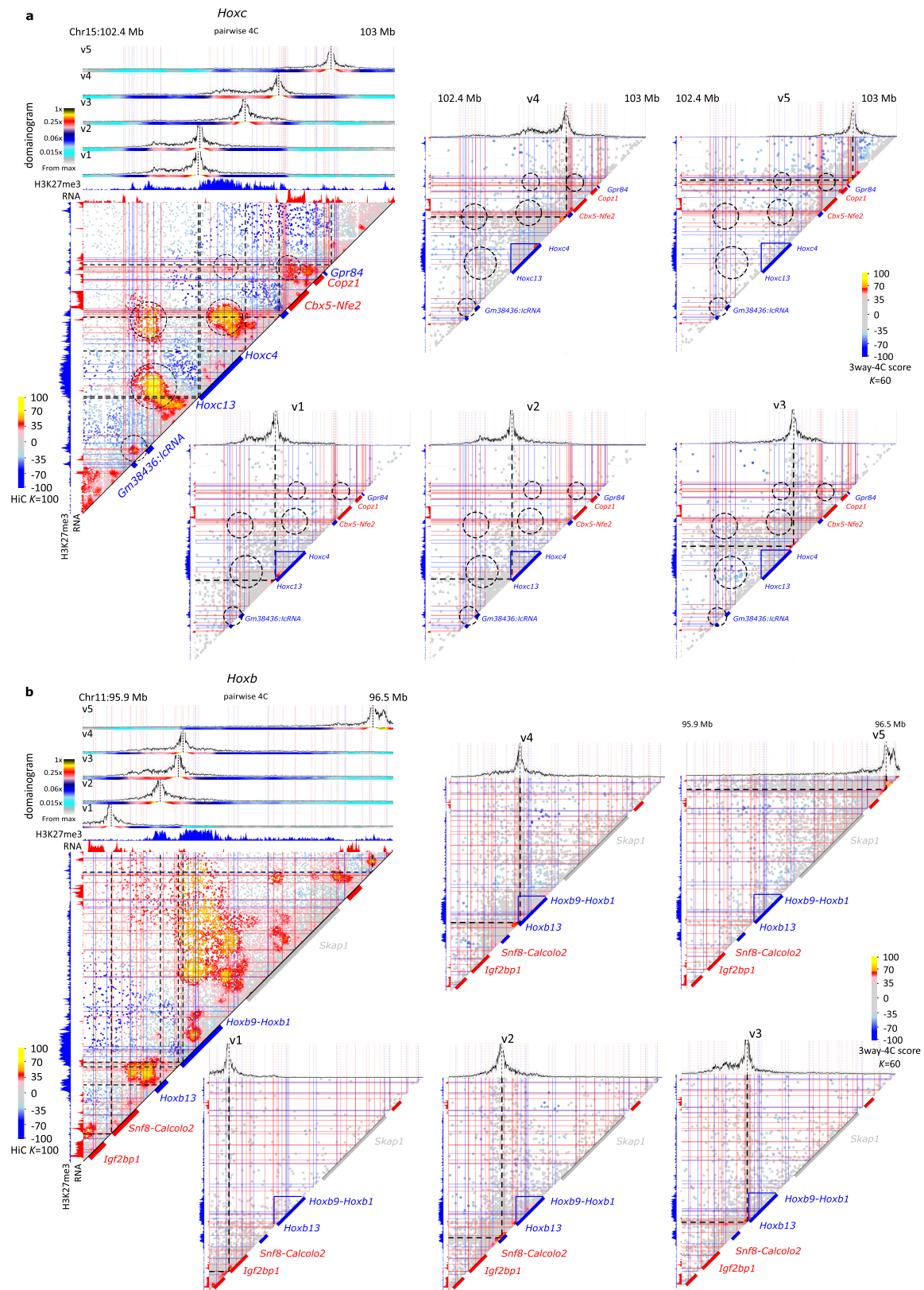


Extended Data Figure 8 | Hi-C and 3way-4C analyses for the β -globin cluster. Similar to Fig. 3a, b. Multiple strong loops^{32,33} are shown by Hi-C, marked H1 to H7. 3-Walk distribution suggests enrichment of multi-contacts between the locus control region (LCR) and the globin genes from the v4, v5 and v6 viewpoints (located at or near the LCR). In addition, the olfactory receptor (OR) cluster upstream the globin

cluster shows weak expression (marked pink) as well as involvement in pairwise contacts with baits within the LCR and downstream of it. This may suggest enrichment of local contacts between the β -globin regulatory machinery and that sites within this cluster are involved in invasion of gene expression into this repressed TAD.



Extended Data Figure 9 | Hi-C and 3way-4C analyses for the *Hoxa* cluster in mES cells.



Extended Data Figure 10 | Hi-C and 3way-4C analyses for the *Hoxc* and *Hoxb* cluster in mES cells. Annotation is as described for Fig. 4a.

m⁶A potentiates *Sxl* alternative pre-mRNA splicing for robust *Drosophila* sex determination

Irmgard U. Haussmann^{1,2}, Zsuzsanna Bodi^{3*}, Eugenio Sanchez-Moran^{1*}, Nigel P. Mongan^{4*}, Nathan Archer³, Rupert G. Fray³ & Matthias Soller¹

N⁶-methyladenosine (m⁶A) is the most common internal modification of eukaryotic messenger RNA (mRNA) and is decoded by YTH domain proteins^{1–7}. The mammalian mRNA m⁶A methylosome is a complex of nuclear proteins that includes METTL3 (methyltransferase-like 3), METTL14, WTAP (Wilms tumour 1-associated protein) and KIAA1429. *Drosophila* has corresponding homologues named Ime4 and KAR4 (Inducer of meiosis 4 and Karyogamy protein 4), and Female-lethal (2)d (Fl(2)d) and Virilizer (Vir)^{8–12}. In *Drosophila*, *fl(2)d* and *vir* are required for sex-dependent regulation of alternative splicing of the sex determination factor *Sex lethal* (*Sxl*)¹³. However, the functions of m⁶A in introns in the regulation of alternative splicing remain uncertain³. Here we show that m⁶A is absent in the mRNA of *Drosophila* lacking *Ime4*. In contrast to mouse and plant knockout models^{5,7,14}, *Drosophila Ime4*-null mutants remain viable, though flightless, and show a sex bias towards maleness. This is because m⁶A is required for female-specific alternative splicing of *Sxl*, which determines female physiognomy, but also translationally represses *male-specific lethal 2* (*msl-2*) to prevent dosage compensation in females. We further show that the m⁶A reader protein YT521-B decodes m⁶A in the sex-specifically spliced intron of *Sxl*, as its absence phenocopies *Ime4* mutants. Loss of m⁶A also affects alternative splicing of additional genes, predominantly in the 5' untranslated region, and has global effects on the expression of metabolic genes. The requirement of m⁶A and its reader YT521-B for female-specific *Sxl* alternative splicing reveals that this hitherto enigmatic mRNA modification constitutes an ancient and specific mechanism to adjust levels of gene expression.

In mature mRNA the m⁶A modification is most prevalently found around the stop codon as well as in 5' untranslated regions (UTRs) and in long exons in mammals, plants and yeast^{2,3,6,7,15}. Since methylosome components predominantly localize to the nucleus, it has been speculated that m⁶A localized in pre-mRNA introns could have a role in alternative splicing regulation in addition to such a role when present in long exons^{9–12,16}. This prompted us to investigate whether m⁶A is required for *Sxl* alternative splicing, which determines female sex and prevents dosage compensation in females¹³. We generated a null allele of the *Drosophila* METTL3 methyltransferase homologue *Ime4* by imprecise excision of a *P* element inserted in the promoter region. The excision allele $\Delta 22-3$ deletes most of the protein-coding region, including the catalytic domain, and is thus referred to as *Ime4*^{null} (Fig. 1a). These flies are viable and fertile, but both flightless and this phenotype can be rescued by a genomic construct restoring *Ime4* (Fig. 1a, b). *Ime4* shows increased expression in the brain and, as in mammals and plants¹⁷, localizes to the nucleus (Fig. 1c, d).

Following RNase T1 digestion and ³²P end-labelling of RNA fragments, we detected m⁶A after guanosine (G) in poly(A) mRNA of adult flies at relatively low levels compared to other eukaryotes

(m⁶A/A ratio: 0.06%, Fig. 1g)^{2,3,5}, but at higher levels in unfertilized eggs (0.18%, Extended Data Fig. 1). After enrichment with an anti-m⁶A antibody, m⁶A is readily detected in poly(A) mRNA, but absent from *Ime4*^{null} flies (Fig. 1h–j).

As found in other systems, and consistent with a potential role in translational regulation^{18–21}, m⁶A was detected in polysomal mRNA (0.1%, Fig. 1k), but not in the poly(A)-depleted rRNA fraction. This also confirmed that any m⁶A modification in rRNA is not after G in *Drosophila* (Fig. 1l).

Consistent with our hypothesis that m⁶A plays a role in sex determination and dosage compensation, the number of *Ime4*^{null} females was reduced to 60% compared to the number of males ($P < 0.0001$), whereas in the control strain female viability was 89% (Fig. 2a). The key regulator of sex determination in *Drosophila* is the RNA-binding protein *Sxl*, which is specifically expressed in females. *Sxl* positively auto-regulates expression of itself and its target *transformer* (*tra*) through alternative splicing to direct female differentiation¹³. In addition, *Sxl* suppresses translation of *msl-2* to prevent upregulation of transcription on the X chromosome for dosage compensation (Fig. 2b); full suppression also requires maternal factors²². Accordingly, female viability was reduced to 13% by removal of maternal m⁶A together with zygotic heterozygosity for *Sxl* and *Ime4* (*Ime4* $\Delta 22-3$ females crossed with *Sxl*^{7B0} males, a *Sxl* null allele, $P < 0.0001$). Female viability of this genotype is completely rescued by a genomic construct (Fig. 2a) or by preventing ectopic activation of dosage compensation by removal of *msl-2* (*msl-2*²²⁷/Df(2L)Exel7016, Fig. 2a). Hence, females are non-viable owing to insufficient suppression of *msl-2* expression, resulting in upregulation of gene expression on the X chromosome from reduced *Sxl* levels. In the absence of *msl-2*, disruption of *Sxl* alternative splicing resulted in females with sexual transformations (32%, $n = 52$) displaying male-specific features such as sex combs (Fig. 2c–e), which were mosaic to various degrees, indicating that *Sxl* threshold levels are affected early during establishment of sexual identities of cells and/or their lineages¹³. In the presence of maternal *Ime4*, *Sxl* and *Ime4* do not genetically interact (*Sxl*^{7B0}/FM7 females crossed with *Ime4*^{null} males, 103% female viability, $n = 118$). In addition, *Sxl* is required for germline differentiation in females and its absence results in tumorous ovaries²³. Consistent with this, we detected tumorous ovaries in *Sxl*^{7B0}/+;*Ime4*^{null}/+ daughters from *Ime4*^{null} females (22%, $n = 18$, Extended Data Fig. 2), but not in homozygous *Ime4*^{null} or heterozygous *Sxl*^{7B0} females ($n = 20$ each).

Furthermore, levels of the *Sxl* female-specific splice form were reduced to approximately 50%, consistent with a role for m⁶A in *Sxl* alternative splicing (Fig. 2f and Extended Data Fig. 3a). As a result, female-specific splice forms of *tra* and *msl-2* were also significantly reduced in adult females (Fig. 2f and Extended Data Fig. 3b, c).

To obtain more comprehensive insights into *Sxl* alternative splicing defects in *Ime4*^{null} females, we examined splice junction reads from

¹School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. ²School of Life Science, Faculty of Health and Life Sciences, Coventry University, Coventry CV1 5FB, UK. ³School of Biosciences, Plant Science Division, University of Nottingham, Sutton Bonington, Loughborough LE12 5RD, UK. ⁴School of Veterinary Medicine and Sciences, University of Nottingham, Sutton Bonington, Loughborough LE12 5RD, UK.

*These authors contributed equally to this work.

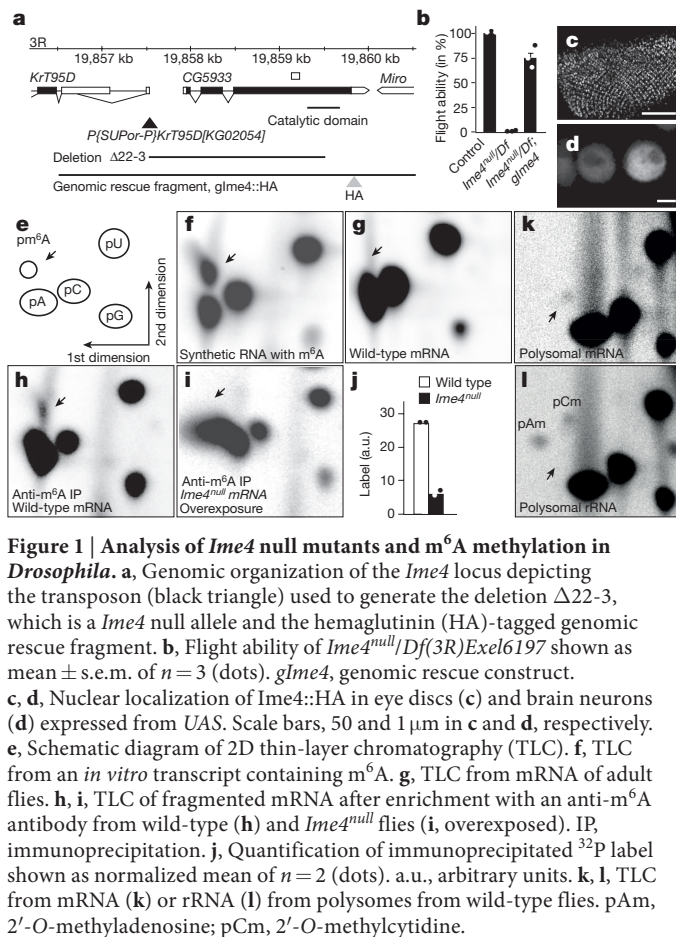


Figure 1 | Analysis of *Ime4* null mutants and m^6A methylation in *Drosophila*. **a**, Genomic organization of the *Ime4* locus depicting the transposon (black triangle) used to generate the deletion $\Delta 22-3$, which is a *Ime4* null allele and the hemagglutinin (HA)-tagged genomic rescue fragment. **b**, Flight ability of *Ime4*^{null}/Df(3R)Exel6197 shown as mean \pm s.e.m. of $n = 3$ (dots). *glme4*, genomic rescue construct. **c**, **d**, Nuclear localization of *Ime4*::HA in eye discs (**c**) and brain neurons (**d**) expressed from *UAS*. Scale bars, 50 and 1 μ m in **c** and **d**, respectively. **e**, Schematic diagram of 2D thin-layer chromatography (TLC). **f**, TLC from an *in vitro* transcript containing m^6A . **g**, TLC from mRNA of adult flies. **h**, **i**, TLC of fragmented mRNA after enrichment with an anti- m^6A antibody from wild-type (**h**) and *Ime4*^{null} flies (**i**, overexposed). IP, immunoprecipitation. **j**, Quantification of immunoprecipitated ^{32}P label shown as normalized mean of $n = 2$ (dots). a.u., arbitrary units. **k**, **l**, TLC from mRNA (**k**) or rRNA (**l**) from polysomes from wild-type flies. pAm, 2'-O-methyladenosine; pCm, 2'-O-methylcytidine.

RNA-seq. Besides the significant increase in inclusion of the male-specific *Sxl* exon in *Ime4*^{null} females (Fig. 2f–h and Extended Data Fig. 3a), cryptic splice sites and increased numbers of intronic reads were detected in the regulated intron. Consistent with our reverse

transcription polymerase chain reaction (RT-PCR) analysis of *tra*, the reduction of female splicing in the RNA sequencing is modest, and as a consequence, alternative splicing differences of *Tra* targets *dsx* and *fru* were not detected in whole flies, suggesting that cell-type-specific fine-tuning is required to generate splicing robustness rather than being an obligatory regulator (Extended Data Fig. 4a–c). In agreement with dosage-compensation defects as a main consequence of *Sxl* dysregulation in *Ime4*^{null} mutants, X-linked, but not autosomal, genes are significantly upregulated in *Ime4*^{null} females compared to controls ($P < 0.0001$, Extended Data Fig. 4d, e).

Furthermore, *Sxl* mRNA is enriched in pull-downs with an m^6A antibody compared to m^6A -deficient mRNA added for quantification (Fig. 2i). This enrichment is comparable to what was observed for m^6A -pull-down from yeast mRNA²⁴.

To map m^6A sites in the intron of *Sxl*, we employed an *in vitro* m^6A methylation assay using *Drosophila* nuclear extracts and labelled substrate RNA. m^6A methylation activity was detected in the vicinity of alternatively spliced exons (Fig. 2j, RNAs B, C, and E). Further fine-mapping localized m^6A in RNAs C and E to the proximity of *Sxl*-binding sites (Extended Data Fig. 5). Likewise, the female-lethal single amino acid substitution alleles *fl(2)d*¹ and *vir*^{2F} interfere with *Sxl* recruitment, resulting in impaired *Sxl* auto-regulation and inclusion of the male-specific exon²⁵. Female lethality of these alleles can be rescued by *Ime4*^{null} heterozygosity ($P < 0.0001$, Fig. 2k), further demonstrating the involvement of the m^6A methylome in *Sxl* alternative splicing.

Next, we globally analysed alternative splicing changes in *Ime4*^{null} females compared to the wild-type control strain. As described earlier (Fig. 2h), a statistically significant reduction in female-specific alternative splicing of *Sxl* (Δ PSI (difference in percentage spliced in) = 0.34, $q = 9 \times 10^{-8}$) was observed. In addition, 243 alternative splicing events in 163 genes were significantly different in *Ime4*^{null} females ($q < 0.05$, Δ PSI > 0.2), equivalent to around 2% of alternatively spliced genes in *Drosophila* (Supplementary Table 1). Six genes for which the alternative splicing products could be distinguished on agarose gels were confirmed by RT-PCR (Extended Data Fig. 6). Notably, lack of *Ime4* did not affect global alternative splicing and no specific type of alternative splicing event was preferentially affected. However, alternative first exon (18% versus 33%) and

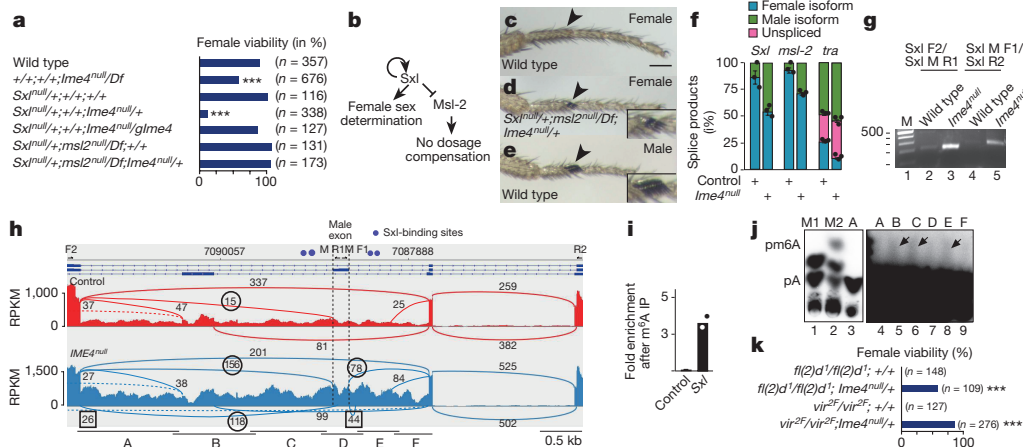


Figure 2 | m^6A methylation is required for *Sxl* alternative splicing in sex determination and dosage compensation. **a**, Female viability of indicated genotypes devoid of maternal m^6A (n , total number of flies, *** $P < 0.0001$). **b**, Schematic depicting *Sxl* control of female differentiation. **c–e**, Front legs of indicated genotypes. Scale bar, 100 μ m. The arrowhead points towards the position of the sex comb normally present only in males (magnified in insets). **f**, Ratio of sex-specific splice isoforms from adult females from RT-PCR shown as mean \pm s.e.m. ($n = 3$, $P < 0.01$ for the change in female isoforms). **g**, RT-PCR for male-specific *Sxl* splicing in control and *Ime4*^{null} females. Lanes are numbered at the bottom. M, DNA size marker. **h**, Sashimi plot depicting Tophat-mapped RNA sequencing reads and exon junction reads

from control and *Ime4*^{null} females below the annotated gene model. Male-specific splice junction reads are circled and cryptic splice sites are boxed. RNA fragments used for m^6A *in vitro* methylation assays are indicated at the bottom (A–F). Primers are indicated on top of exons with arrows. **i**, Presence of m^6A in *Sxl* transcripts detected by m^6A immunoprecipitation followed by qPCR from nuclear mRNA of early embryos (shown as mean of $n = 2$, dots). **j**, One-dimensional TLC of *in vitro*-methylated, [^{32}P]-ATP-labelled substrate RNAs shown in **g**. Nucleotide markers from *in vitro* transcripts in the absence (M1) or presence (M2) of m^6A . The right image shows an overexposure of the same TLC, lanes 3 and 4 show the same sample. **k**, Rescue of female lethality of female-lethal *fl(2)d*¹ and *vir*^{2F} alleles by removal of one copy of *Ime4*.

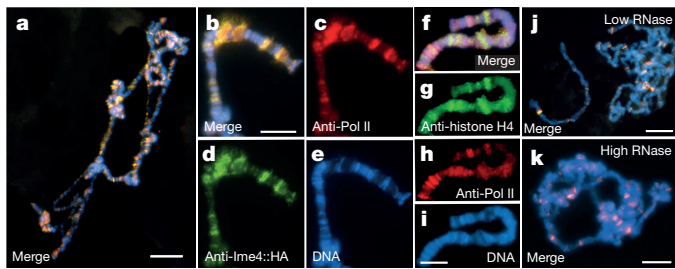


Figure 3 | Ime4 co-localizes to sites of transcription. **a–e**, Polytene chromosomes from salivary glands expressing IME::HA stained with anti-Pol II (red, **c**), anti-HA (green, **d**) and DAPI (DNA, blue, **e**), or merged (yellow, **a** and **b**). **f–i**, Polytene chromosomes stained with anti-Pol II (red, **h**), anti-histone H4 (green, **g**) and DAPI (DNA, blue, **i**), or merged (yellow, **f**). **j, k**, Polytene chromosomes treated with low (**j**, $2\mu\text{g ml}^{-1}$) and high (**k**, $10\mu\text{g ml}^{-1}$) RNase A concentration before staining with anti-Pol II, anti-histone H4 and DAPI. Scale bars, $20\mu\text{m}$ (**a, j, k**) and $5\mu\text{m}$ (**e, i**).

mutually exclusive exon (2% versus 15%) events were reduced in *Ime4^{null}* compared to a global breakdown of alternative splicing in wild-type *Drosophila*, mostly to the extent of retained introns (16% versus 6%), alternative donor (16% versus 9%) and unclassified events (14% versus 6%) (Extended Data Fig. 7a). Notably, the majority of affected alternative splicing events in *Ime4^{null}* were located to the 5' UTR, and these genes had a significantly higher number of AUG start codons in their 5' UTR compared to the 5' UTRs of all genes (Extended Data Fig. 7b, c). Such a feature has been shown to be relevant to translational control under stress conditions²⁶.

The majority of the 163 differentially alternatively spliced genes in *Ime4* females are broadly expressed (59%), while most of the remainder are expressed in the nervous system (33%), consistent with higher expression of *Ime4* in this tissue (Extended Data Fig. 7d). Accordingly, Gene Ontology analysis revealed a highly significant enrichment for

genes involved synaptic transmission ($P < 7 \times 10^7$, Supplementary Table 1).

Since the absence of *m*⁶A affects alternative splicing, *m*⁶A marks are probably deposited co-transcriptionally before splicing. Co-staining of polytene chromosomes with antibodies against haemagglutinin (HA)-tagged *Ime4* and RNA Pol II revealed broad co-localization of *Ime4* with sites of transcription (Fig. 3a–e), but not with condensed chromatin—visualized with antibodies against histone H4 (Fig. 3f–i). Furthermore, localization of *Ime4* to sites of transcription is RNA-dependent, as staining for *Ime4*, but not for RNA Pol II, was reduced in an RNase-dependent manner (Fig. 3j, k).

Although *m*⁶A levels after G are low in *Drosophila* compared to other eukaryotes, broad co-localization of *Ime4* to sites of transcription suggests profound effects on the gene expression landscape. Indeed, differential gene expression analysis revealed 408 differentially expressed genes (≥ 2 -fold change, $q \leq 0.01$) where 234 genes were significantly upregulated and 174 significantly downregulated in neuron-enriched head/thorax of adult *Ime4^{null}* females ($q < 0.01$, at least twofold, Supplementary Table 2). Cataloguing these genes according to function reveals prominent effects on gene networks involved in metabolism, including reduced expression of 17 genes involved in oxidative phosphorylation ($P < 0.0001$, Supplementary Table 2). Notably, overexpression of the *m*⁶A mRNA demethylase FTO in mice leads to an imbalance in energy metabolism resulting in obesity²⁷.

Next, we tested whether either of the two substantially divergent YTH proteins, YT521-B and CG6422 (Fig. 4a), decodes *m*⁶A marks in *Sxl* mRNA. When transiently transfected into male S2 cells, YT521-B localizes to the nucleus, whereas CG6422 is cytoplasmic (Fig. 4b–d, Extended Data Fig. 8). Nuclear YT521-B can switch *Sxl* alternative splicing to the female mode and also binds to the *Sxl* intron in S2 cells (Fig. 4e, f). *In vitro* binding assays with the YTH domain of YT521-B demonstrate increased binding of *m*⁶A-containing RNA (Extended Data Fig. 9). *In vivo*, YT521-B also localizes to the sites of transcription (Extended Data Fig. 10).

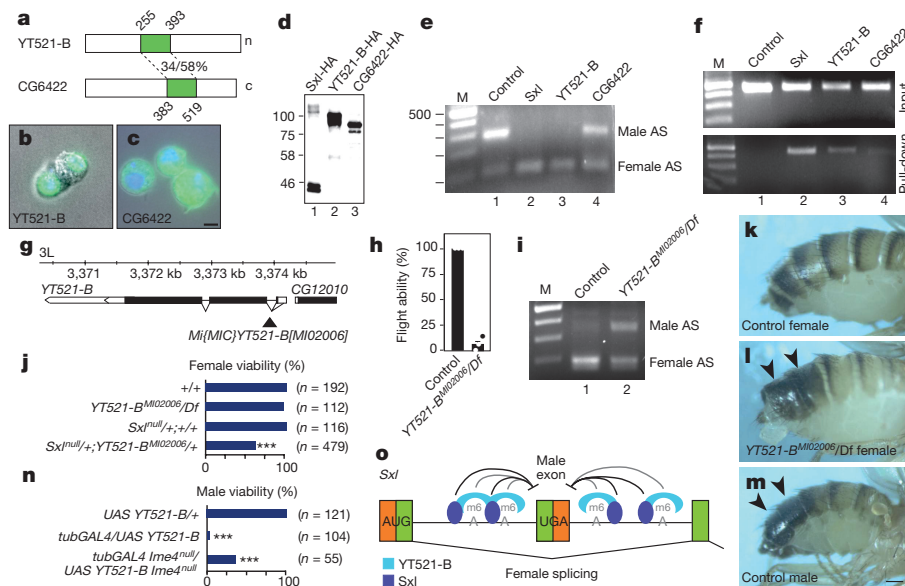


Figure 4 | YTH protein YT521-B decodes *m*⁶A methylation in *Sxl*. **a**, Domain organization of *Drosophila* YTH proteins (YTH domain in green). n, nuclear; c, cytoplasmic. **b–d**, Cellular localization and size of HA-tagged YT521-B and CG6422 in S2 cells. Scale bar, $1\mu\text{m}$. **e**, Suppression of male-specific *Sxl* alternative splicing (AS) upon expression of *Sxl* and YT521-B, but not CG6422 in male S2 cells. **f**, Binding of YT521-B to pre-mRNA of the regulated *Sxl* intron. **g**, Genomic organization of the *YT521-B* locus depicting the transposon (black triangle) disrupting the ORF. **h**, Flight ability of *YT521-B^{M102006}/Df(3L)Exel6094* shown as mean \pm s.e.m. ($n = 3$). **i**, *Sxl* alternative splicing

in female wild-type and *YT521-B^{M102006}/Df(3L)Exel6094* flies. **j**, Female viability of indicated genotypes (n , total number of flies) reared at 29°C . **k–m**, Abdominal pigmentation of indicated genotypes reared at 29°C . The arrowheads point towards the position of the dark pigmentation normally present only in males. Scale bar, $100\mu\text{m}$. **n**, YT521-B was overexpressed from a UAS transgene with *tubulinGAL4* (2nd chromosome insert) in wild-type or *Ime4^{null}* flies at 27°C . **o**, Model for female-specific *Sxl* alternative splicing by *Sxl*, *m*⁶A and its reader YT521-B in co-operatively suppressing inclusion of the male-specific exon.

To further examine the role of YT521-B in decoding m⁶A we analysed *Drosophila* strain YT521-B^{M102006}, where a transposon in the first intron disrupts YT521-B. This allele is also viable (YT521-B^{M102006}/Df(3L)Exel6094; Fig. 4g, h, j), and phenocopies the flightless phenotype and the female Sxl splicing defect of *Ime4*^{null} flies (Fig. 4h, i). Likewise, removal of maternal YT521-B together with zygotic heterozygosity for Sxl and YT521-B reduces female viability ($P < 0.0001$, Fig. 4j) and results in sexual transformations (57%, $n = 32$) such as male abdominal pigmentation (Fig. 4k–m). In addition, overexpression of YT521-B results in male lethality, which can be rescued by removal of *Ime4*, further reiterating the role of m⁶A in Sxl alternative splicing ($P < 0.0001$, Fig. 4n). Since YT521-B phenocopies *Ime4* for Sxl splicing regulation, it is the main nuclear factor for decoding m⁶A present in the proximity of the Sxl-binding sites. YT521-B bound to m⁶A assists Sxl in repressing inclusion of the male-specific exon, thus providing robustness to this vital gene regulatory switch (Fig. 4o).

Nuclear localization of m⁶A methylome components suggested a role for this “fifth” nucleotide in alternative splicing regulation. Our discovery of the requirement of m⁶A and its reader YT521-B for female-specific Sxl alternative splicing has important implications for understanding the fundamental biological function of this enigmatic mRNA modification. Its key role in providing robustness to Sxl alternative splicing to prevent ectopic dosage compensation and female lethality, together with localization of the core methylome component Ime4 to sites of transcription, indicates that the m⁶A modification is part of an ancient, yet unexplored mechanism to adjust gene expression. Hence, the recently reported role of m⁶A methylome components in human dosage compensation^{28,29} further support such a role and suggests that m⁶A-mediated adjustment of gene expression might be a key step to allow for the development of the diverse sex determination mechanisms found in nature.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 March; accepted 25 October 2016.

Published online 30 November 2016.

1. Luo, S. & Tong, L. Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. *Proc. Natl Acad. Sci. USA* **111**, 13834–13839 (2014).
2. Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
3. Dominissini, D. *et al.* Topology of the human and mouse m⁶A RNA methylomes revealed by m⁶A-seq. *Nature* **485**, 201–206 (2012).
4. Perry, R. P. & Kelley, D. E. Existence of methylated messenger RNA in mouse L cells. *Cell* **1**, 37–42 (1974).
5. Zhong, S. *et al.* MTA is an *Arabidopsis* messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* **20**, 1278–1288 (2008).
6. Schwartz, S. *et al.* High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* **155**, 1409–1421 (2013).
7. Ke, S. *et al.* A majority of m⁶A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes Dev.* **29**, 2037–2053 (2015).
8. Liu, J. *et al.* A METTL3-METTL14 complex mediates mammalian nuclear RNA N⁶-adenosine methylation. *Nat. Chem. Biol.* **10**, 93–95 (2014).
9. Horiuchi, K. *et al.* Identification of Wilms' tumor 1-associating protein complex and its role in alternative splicing and the cell cycle. *J. Biol. Chem.* **288**, 33292–33302 (2013).

10. Bokar, J. A., Shambaugh, M. E., Polayes, D., Matera, A. G. & Rottman, F. M. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N⁶-adenosine)-methyltransferase. *RNA* **3**, 1233–1247 (1997).
11. Penálva, L. O. *et al.* The *Drosophila* *tl(2)d* gene, required for female-specific splicing of Sxl and tra pre-mRNAs, encodes a novel nuclear protein with a HQ-rich domain. *Genetics* **155**, 129–139 (2000).
12. Niessen, M., Schreiner, R. & Nothiger, R. Molecular identification of virilizer, a gene required for the expression of the sex-determining gene Sex-lethal in *Drosophila melanogaster*. *Genetics* **157**, 679–688 (2001).
13. Schütt, C. & Nothiger, R. Structure, function and evolution of sex-determining systems in Dipteran insects. *Development* **127**, 667–677 (2000).
14. Geula, S. *et al.* Stem cells. m⁶A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347**, 1002–1006 (2015).
15. Luo, G. Z. *et al.* Unique features of the m⁶A methylome in *Arabidopsis thaliana*. *Nat. Commun.* **5**, 5630 (2014).
16. Xiao, W. *et al.* Nuclear m⁶A reader YTHDC1 regulates mRNA splicing. *Mol. Cell* **61**, 507–519 (2016).
17. Hongay, C. F. & Orr-Weaver, T. L. *Drosophila* Inducer of Meiosis 4 (IME4) is required for Notch signaling during oogenesis. *Proc. Natl Acad. Sci. USA* **108**, 14855–14860 (2011).
18. Bodi, Z., Bottley, A., Archer, N., May, S. T. & Fray, R. G. Yeast m⁶A methylated mRNAs are enriched on translating ribosomes during meiosis, and under rapamycin treatment. *PLoS One* **10**, e0132090 (2015).
19. Wang, X. *et al.* N⁶-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161**, 1388–1399 (2015).
20. Meyer, K. D. *et al.* 5' UTR m⁶A Promotes Cap-Independent Translation. *Cell* **163**, 999–1010 (2015).
21. Zhou, J. *et al.* Dynamic m⁶A mRNA methylation directs translational control of heat shock response. *Nature* **526**, 591–594 (2015).
22. Zaharieva, E., Haussmann, I. U., Bräuer, U. & Soller, M. Concentration and localization of co-expressed ELAV/Hu proteins control specificity of mRNA processing. *Mol. Cell. Biol.* **35**, 3104–3115 (2015).
23. Salz, H. K. Sex, stem cells and tumors in the *Drosophila* ovary. *Fly (Austin)* **7**, 3–7 (2013).
24. Bodi, Z., Button, J. D., Grierson, D. & Fray, R. G. Yeast targets for mRNA methylation. *Nucleic Acids Res.* **38**, 5327–5335 (2010).
25. Hilfiker, A., Amrein, H., Dübendorfer, A., Schreiner, R. & Nothiger, R. The gene virilizer is required for female-specific splicing controlled by Sxl, the master gene for sexual development in *Drosophila*. *Development* **121**, 4017–4026 (1995).
26. Starck, S. R. *et al.* Translation from the 5' untranslated region shapes the integrated stress response. *Science* **351**, aad3867 (2016).
27. Church, C. *et al.* Overexpression of Fto leads to increased food intake and results in obesity. *Nat. Genet.* **42**, 1086–1092 (2010).
28. Moindrot, B. *et al.* A pooled shRNA screen identifies Rbm15, Spen, and Wtap as factors required for Xist RNA-mediated silencing. *Cell Reports* **12**, 562–572 (2015).
29. Patil, D. P. *et al.* m⁶A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Horabin, N. Perrimon and the Bloomington, Harvard and Kyoto stock centres for fly lines, BacPac for DNA clones, E. Zaharieva and M. L. Li for help with imaging, W. Arlt and R. Michell for comments on the manuscript, and J.-Y. Roignant for communication of results before publication. We acknowledge funding from the BBSRC (BB/M008606/1) to R.F.

Author Contributions I.U.H. and M.S. performed biochemistry, cell biology and genetic experiments, E.S.M. stained chromosomes, and Z.B., N.A. and R.F. performed biochemistry experiments. N.M. analysed sequencing data. I.U.H., R.F. and M.S. conceived the project and wrote the manuscript with help from N.M. and Z.B.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S. (m.soller@bham.ac.uk).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Drosophila genetics, generation of constructs and transgenic lines. The deletion allele *Ime4*^{Δ22-3} was obtained from imprecise excision of the transposon *P(SUPor-P)KrT95D* and mapped by primers 5933 F1 (CTCGCTCTA TTTCTCTTCAGCACTCG) and 5933 R9 (CCTCCGCAACGATCACAT CGCAATCGAG). To obtain a viable line of *Ime4*^{null}, the genetic background was cleaned by out-crossing to *Df(3R)Exel6197*. Flight ability was scored as the number of flies capable of flying out of a Petri dish within 30 s for groups of 15–20 flies for indicated genotypes. Viability was calculated from the numbers of females compared to males of the correct genotype and statistical significance was determined by a χ^2 test (GraphPad Prism). Unfertilized eggs were generated by expressing sex-peptide in virgin females as described³⁰.

The genomic rescue construct was retrieved by recombineering (GeneBridges) from BAC clone CH321-79E18 by first cloning homology arms with *SpeI* and *Aac65I* into *pUC3GLA* separated by an *EcoRV* site for linearization (CTCCGCCGCCG AACC GCCGCTCTCTCCGCCACTTTG CAGGTTGAGCGGACCGCTCCAG GGCCGCTGCCGCCGCTGCCGCTGATATCCAGCATGGTAGCTGCGGCC ACTCCTAGTCCCGCTTTAACCACAGCTTGGGGTCTCCGTCATCAGG CCGAATTGCTCTGAG). An HA-tag was then fused to the end of the ORF using two PCR amplicons and *SacI* and *XhoI* restriction sites. This construct was the inserted into *PBac{y+attB-3B}VK00002* at 76A as described³¹.

The *Ime4* UAS construct was generated by cloning the ORF from fly cDNA into a modified *pUAST* with primers Adh dMT-A70 F1 EI (GCAGAATTCGAG ATCTAAAGAGCGCTGCTAAAGCAAAAAGAGTCACCATGGCAGATGCGT GGGACATAAAATCAC) and dMT-A70 HA R1 Spe (GGTAACTAGTCTTTTG TATTCATTGATCGACGCCGATTGG) by adding a translation initiation site from the *Adh* gene and two copies of an HA tag to the end of the ORF. This construct was then also inserted into *PBac{y+attB-3B}VK00002* at 76A.

For transient transfection in S2 cells, *YT52B-1* and *CG6422* ORFs were amplified from fly cDNA by a combination of nested and fusion PCR incorporating a translation initiation site from the *Adh* gene using primers CG6422 Adh F1 (GCCTGCTAAAGCAAAAAGAGTCACCATGTCAGGCGTG TCCAGATGAAAATACCAG), pACT Adh CG6422 F1 (CCAGAGACCCCGGA TCCAGATATCAAAAGAGCGCTGCTAAAGCAAAAAGAGTCACCATC), CG 6422 Adh R1, (GATTCCTGCGAACAGGTCCCGTGGCGCAAAC) and CG6422 3' F1 (CCCACGGGACCTGTTCGACGAATCTAG), CG6422 3' R1 (CATTGC TTCGATTTTATCCTTGTCCTGTCTTAAAGCGCAGCCGATTTTAAT TTG), pACT CG6422 3×HA R1 (GTGGAGATCCATGGTGGCGGAGCTCGA GGAATATTCATTGCTTCGCTTTTATCTTGTGTC) for CG6422 and primers YT521 Adh F1, (AAGCAAAAAGAGTCACATGCCAAGAGCAGCCGTA AACAAACGCTGCCGATGCGCGAG), pACT Adh YT521 F1 (CCAGAGACC CCGGATCCAGATATCAAGAGCGCTGCTAAAGCAAAAAGAGTCACAT GCC), YT521 Adh R1 (TGCCATCCGGGCGAATCCTGCAAAATTTACC ACTCTCGTTGACCGGAAAGTACGAGGAC) and YT521 3' F1 (GC AGGATTCGCCCCGATGGCAGCCCCCTCAC), pACT YT521 R1 (GGTGGAG ATCCATGGTGGCGGAGCTCGAGCGCTGTGTCCCGATAGCTTCGCTG) for *YT521-B*, and cloned into a modified *pACT* using Gibson Assembly (NEB) also incorporating HA epitope tags at the C terminus. Constructs were verified by Sanger sequencing. The *Sxl*-HA expression vector was a gift from N. Perrimon³².

The *YT521-B* UAS construct was generated by sub-cloning the ORF from the pACT vector into a modified *pUAST* with primers YT521 Adh F1 (AAGCAAAA AAGAGTCACATGCCAAGAGCAGCCGTAACAAACGCTGCCGATGCG CGAG), YT521 Adh F2 (TAGGGAATTGGGAATTCGAGATCTAAAGAGCCT GCTAAAGCAAAAAGAGTCACATGCC) and YT521 3' R1 (GGGCACGT CGTAGGGGTACAGACTAGTCTCGAGGCGCCTGTGTCCCGATAGCTTC GCTG) by adding a translation initiation site from the *Adh* gene and two copies of an HA tag to the end of the ORF. This construct was then also inserted into *PBac{y+attB-3B}VK00002* at 76A.

Essential parts of all DNA constructs were sequence-verified.

Cell culture, transfections and immune-staining of S2 cells. S2 cells (ATCC) were cultured in Insect Express medium (Lonza) with 10% heat-inactivated FBS and 1% penicillin/streptomycin. The *Drosophila* S2 cell line was verified to be male by analysing *Sxl* alternative splicing using species-specific primers *Sxl* F2 (ATGTACGGCAACAATAATCCGGGTAG) and *Sxl* R2 (CATTGTAACACGACGCGACGATG) to confirm species and gender (Extended Data Fig. 8). Transient transfections were done with Mirus Reagent (Bioline) according to the manufacturer's instruction and cells were assayed 48 h after transfection for protein expression or RNA binding of expressed proteins. To adhere S2 cells to a solid support, Concanavalin A (Sigma) coated glass slides

(in 0.5 mg ml⁻¹) were added 1 day before transfection, and cells were stained 48 h after transfection with antibodies as described. Transfections and follow up experiments were repeated at least once.

RNA extraction, RT-PCR, qPCR, immunoprecipitation and western blots. Total RNA was extracted using Tri-reagent (SIGMA) and reverse transcription was done with Superscript II (Invitrogen) according to the manufacturer's instructions using an oligodT17V primer. PCR for *Sxl*, *tra*, *msl2* and *ewg* was done for 30 cycles with 1 µl of cDNA with primers *Sxl* F2, *Sxl* R2 or *Sxl* NP R3 (GAGAATGGGACATCCCAATCCACG), *Sxl* M F1 (GCCAGAGA AAGAAGCAGCCACCATTATCAC), *Sxl* M R1 (GCGTTTCGTTGGCGAG GAGACCATTGGG), *Tra* FOR (GGATGCCGACAGCAGTGGAAAC), *Tra* REV (GATCTGGAGCGAGTGCCTCTG), *Ms1-2* F1 (CACTGCGGTCA CACTGGCTTCGCTCAG), *Ms1-2* R1 (CTCCTGGGCTAGTTACCTGCAATTC CTC), *Ewg* 4F and *Ewg* 5R and quantified with ImageQuant (BioRad)²². Experiments included at least three biological replicates.

For qPCR, reverse transcription was carried out on input and pull-down samples spiked with yeast RNA using ProtoScript II reverse transcriptase and random nanomers (NEB). Quantitative PCR was carried out using 2× SensiMix Plus SYBR Low ROX master mix (Quantace) using normalizer primers ACT1 F1 (TTAC GTGCCTTGGACTTCG) and ACT1 R1 (TACCGGCAGATTCCAAACCC) and for *Sxl*, *Sxl* ZB F1 (CACCACAATGGCAGCAGTAG) and *Sxl* ZB R1 (GGGGTT GCTGTTTGTGAGT). Samples were run in triplicate for technical repeats and duplicate for biological repeats. Relative enrichment levels were determined by comparison with yeast ACT1, using the 2^{-ΔΔC_t} method³³.

For immunoprecipitations of *Sxl* RNA bound to *Sxl* or YTH proteins, S2 cells were fixed in PBS containing 1% formaldehyde for 15 min, quenched in 100 mM glycine and disrupted in IP-Buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.5, 1% NP-40, 5% glycerol). After IP with anti-HA beads (Sigma) for 2 h in the presence of Complete Protein Inhibitor (Roche) and 40 U RNase inhibitors (Roche), IP precipitates were processed for *Sxl* RT-PCR using gene-specific RT primer SP NP2 (CATTCGGATGGCAGAGATGGGAC) and PCR primers *Sxl* NP intF (GAGGTCAGTCTAAGTTATATTCG) and *Sxl* NP R3 as described³¹. Western blots were done as described using rat anti-HA (1:50, clone 3F10, Roche) and HRP-coupled secondary goat anti-rat antibodies (Molecular Probes)³⁴. All experiments were repeated at least once from biological samples.

Analysis of m⁶A levels. Poly(A) mRNA from at least two rounds of oligo dT selection was prepared according to the manufacturer (Promega). For each sample, 10–50 ng of mRNA was digested with 1 µl of Ribonuclease T1 (1,000 U µl⁻¹; Fermentas) in a final volume of 10 µl in polynucleotide kinase buffer (PNK, NEB) for 1 h at 37 °C. The 5' end of the T1-digested mRNA fragments were then labelled using 10 U T4 PNK (NEB) and 1 µl [γ -³²P]-ATP (6,000 Ci mmol⁻¹; Perkin-Elmer). The labelled RNA was precipitated, resuspended in 10 µl of 50 mM sodium acetate buffer (pH 5.5), and digested with P1 nuclease (Sigma-Aldrich) for 1 h at 37 °C. Two microlitres of each sample was loaded on cellulose TLC plates (20 × 20 cm; Fluka) and run in a solvent system of isobutyric acid: 0.5 M NH₄OH (5:3, v/v), as the first dimension, and isopropanol:HCl:water (70:15:15, v/v/v), as the second dimension. TLCs were repeated from biological replicates. The identification of the nucleotide spots was carried out using m⁶A-containing synthetic RNA. Quantification of ³²P was done by scintillation counting (Packard Tri-Carb 2300TR). For the quantification of spot intensities on TLCs or gels, a storage phosphor screen (K-Screen; Kodak) and Molecular Imager FX in combination with QuantityOne software (BioRad) were used.

For immunoprecipitation of m⁶A mRNA, poly(A) mRNA was digested with RNase T1 and 5' labelled. The volume was then increased to 500 µl with IP buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.5, 0.05% NP-40). IPs were then done with 2 µl of affinity-purified polyclonal rabbit m⁶A antibody (Synaptic Systems) and protein A/G beads (SantaCruz).

Polysome profiles. Whole-fly extracts were prepared from 20–30 adult *Drosophila* previously frozen in liquid N₂ and ground into fine powder in liquid N₂. Cells were then lysed in 0.5 ml lysis buffer (0.3 M NaCl, 15 mM MgCl₂, 15 mM Tris-HCl pH 7.5, cycloheximide 100 µg ml⁻¹, heparin (sodium salt) 1 mg ml⁻¹, 1% Triton X-100). Lysates were loaded on 12 ml sucrose gradients and spun for 2 h at 38,000 r.p.m. at 4 °C. After the gradient centrifugation 1-ml fractions were collected and precipitated in equal volume of isopropanol. After several washes with 80% ethanol the samples were resuspended in water and processed. Experiments were done in duplicate.

Nuclear extract preparation and in vitro m⁶A methylation essays. *Drosophila* nuclear extracts were prepared from Kc cells as described³⁵. Templates for *in vitro* transcripts were amplified from genomic DNA using the primers listed below and *in vitro* transcribed with T7 polymerase in the presence of [α -³²P]-ATP. DNA templates and free nucleotides were removed by DNase I digestion and Probequant G-50 spin columns (GE Healthcare), respectively. Markers were generated by

using *in vitro* transcripts with or without m^6 ATP (Jena Bioscience), which were then digested with RNase T1, kinased with PNK in the presence of $[\gamma\text{-}^{32}\text{P}]\text{-ATP}$. After phenol extraction and ethanol precipitation, transcripts were digested to single nucleotides with P1 nuclease as above. For *in vitro* methylation, transcripts ($0.5\text{--}1 \times 10^6$ c.p.m.) were incubated for 45 min at 27°C in $10\mu\text{l}$ containing 20 mM potassium glutamate, 2 mM MgCl_2 , 1 mM DTT, 1 mM ATP, 0.5 mM S-adenosylmethionine disulfate tosylate (Abcam), 7.5% PEG 8000, 20 U RNase protector (Roche) and 40% nuclear extract. After phenol extraction and ethanol precipitation, transcripts were digested to single nucleotides with P1 nuclease as above, and then separated on cellulose F TLC plates (Merck) in 70% ethanol, previously soaked in 0.4 M MgSO_4 and dried³⁶. *In vitro* methylation assays were done from biological replicates at least in duplicates.

Primers to amplify parts of the *Sxl* alternatively spliced intron from genomic DNA for *in vitro* transcription with T7 polymerase were *Sxl* A T7 F (GGAGCTAATACGACTCACTATAGGGAGAGGATATGACGGCAACAATAA TCCGGGTAG) and *Sxl* A R (CGCAGACGACGATCAGCTGATTCAAAGTGA AAG), *Sxl* B T7 F (GGAGCTAATACGACTCACTATAGGGAGAGCGCTCG CATTATCCACAGTCGCAC) and *Sxl* B R (GGGTGCCCTCTGTGGCTG CTCTGTTTAC), *Sxl* C T7 F (GGAGCTAATACGACTCACTATAGGGGTCGT ATAATTATGGCACATTATTCAG) and *Sxl* C R (GGGAGTTTGTGTTT TGTATTAGTGGGTG), *Sxl* D T7 F (GGAGCTAATACGACTCACTA TAGGGAGAAAACTCCAGCCCACACACACAC) and *Sxl* D R (GCATATCATATTCGGTTATCATAGTTTAAAG), *Sxl* E T7 F (GGAG CTAATACGACTCACTATAGGGAGAGGGGAAGCAGCTCGTTGTAA AATAC) and *Sxl* E R (GATGTGACGATTTTGCAGTTTCTCGACG), *Sxl* F T7 F (GGAGCTAATACGACTCACTATAGGGAGAGGGGGATCGTT TTGAGGGTCAGTCTAAG) and *Sxl* NP2, *Sxl* C T7 F and *Sxl* C1 R (GTAG TTTTGCTCGGCATTATTGACCTTGAGC), *Sxl* C2 F (GGAGCTAATACG ACTCACTATAGGGAGACTCTCATTTCTATATCCCTGTGCTGACC) and *Sxl* C2 R (CTAATTTCGTGAGCTTGATTTCATTTTGCACAG), *Sxl* C3 F (GGAGCTAATACGACTCACTATAGGGAGACTGTGCAAAATGAAATCAAGC TCACGAAATTAG) and *Sxl* C R, *Sxl* E T7 F and *Sxl* E1 R (AAAAAATCAAA AAAAATACACTTTTGGCATTTCATCAC), *Sxl* E2 F (GGAGCTAATAC GACTCACTATAGGGAGATGAAAAAGTGCCAAAAGTGATTATTTT TTTG), *Sxl* E2 R (AAAAGCATGATGATTTTTTTTTTTTTTGTACTTTTCG AATCACC), *Sxl* E3 F (GGAGCTAATACGACTCACTATAGGGAGAC GGTGATTTCGAAAGTACAAAAAATCAATAC) and *Sxl* E R, *Sxl* C4 F (GAGCTAATACGACTCACTATAGGGAGACTGTGCAAAATGAAATCAAGC AACCGCAAGCAGAGCAGC) and *Sxl* C4 R (GAGTGCCACTTCAAAAT CTCAGATATGC), *Sxl* C5 F (CTAATACGACTCACTATAGGGAGACTCTTT TTTTTTTCTTTTTTACTGTGCAAAATG) and *Sxl* C5 R (AAAAAATAT GCAAAAAAAGTATAGGGCACAAGTTCTCAATTAC), *Sxl* C6 F (GAGCTAATACGACTCACTATAGGGAGACTGTGCAAAATGAAATCAAGC TCACGAAATTAG) and *Sxl* C6 R (CAATTTCATATATGTACGAAA ACAAAGTGAG), *Sxl* E4 F (GGAGCTAATACGACTCACTATAGGGA GAACCAAAATTCGACGTGGGAAGAAAC) and *Sxl* E4 R (TAATCACT TTTGGCACTTTTTCATCATTAAC), *Sxl* E5 F (GGCTAATACGACTCACTAT AGGGAGATTTTTTGTATTTTAAAGTGAAGTGTGCTCC) and *Sxl* E5 R (CACCGAAAAAATAAAAAAATAATCATGGGACTATACTAG), *Sxl* E6 F (GGCTAATACGACTCACTATAGGGAGACTTAAGTGCAATATTTAAAGT GAAACCAATTG) and *Sxl* E6 R (CCCCAGTTATATCAACCGTGAAAT TATGC).

Illumina sequencing and analysis of differential gene expression and alternative splicing. Total RNA was extracted from 15 pulverized head/thoraces previously flash-frozen in liquid nitrogen, using TRIzol reagent from *white* (*w*) control and *w;Ime4 Δ^{22-3}* females that have been outcrossed for several generations to *w; Df(3R)Exel6197* to equilibrate genetic background. Total RNA was treated with DNase I (Ambion) and stranded libraries for Illumina sequencing were prepared after poly(A) selection from total RNA ($1\mu\text{g}$) with the TruSeq stranded mRNA kit (Illumina) using random primers for reverse transcription according to the manufacturer's instructions. Pooled indexed libraries were sequenced on an Illumina HiSeq2500 to yield 40–46 million paired-end 100 bp reads, and in a second experiment 14–19 million single-end 125-bp reads for three controls and mutants each. After demultiplexing, sequence reads were aligned to the *Drosophila* genome (dmel-r6.02) using Tophat2.0.6 (ref. 37). Differential gene expression was determined by Cufflinks-Cuffdiff and the FDR-correction for multiple tests to raw *P* values with $q < 0.05$ considered significant³⁸. alternative splicing was analysed by SPANKI³⁹ and validated for selected genes based on length differences detectable on agarose gels. Illumina sequencing, differential gene expression and alternative splicing analysis was done by Fasteris (Switzerland). For dosage compensation analysis, differential expression analysis of X-linked genes versus autosomal genes in *Ime4 null* mutant was done by filtering Cuffdiff data by a *P* value expression difference significance of $P < 0.05$, which corresponds to a false discovery

rate of 0.167 to detect subtle differences in expression consistent with dosage compensation. Visualization of sequence reads on gene models and splice junctions reads in Sashimi plots was done using Integrated Genome Viewer⁴⁰. For validation of alternative splicing by RT-PCR as described above, the following primers were used: Gprk2 F1 (CCAACCAGCCGAAACTCACAGTGAAGC) and Gprk2 R1 (CAGGGTCTCGGTTTCAGACACAGGCGTC), fl(2)d F1 (GCAGCAAACGA GAAATCAGCTCGCAGCGCAG) and fl(2)d R1 (CACATAGTCTGGAATTCTT GCTCCTTG), A2bp1 F3 (CTGTGGGGCTCAGGGGCATTTTCCCTCCTC) and A2bp1 R1 (CTCCTCTCCCGTGTGTCTTGCCACTCAAC), cv.-c F1 (GGGTT TCCACCTCGACCGGGAAAAAGTCG) and cv.-c R1 (GCGTTTGCGG TTGTGCTCGCGAAGAGAG), CG8312 F1 (GCGCGTGGCCTCCTTCTT ATCGGCACT) and CG8312 R1 (GCGTGGCCACTATAAAGTCCACCTCATC), Chas F2 (CCGATTTCGATTTCGATTTCGATCCTCTCTTC) and Chas R1 (GTCGGTGTCTCGGTGGTGTGGTGGAG). GO enrichment analysis was done with FlyMine. For the analysis of uATGs, an R script was used to count the uATGs in 5' UTRs in all ENSEMBL isoforms of those genes which are differentially spliced in *Ime4* mutants, that were then compared to the mean number of ATGs in all *Drosophila* ENSEMBL 5' UTRs using a *t*-test. Gene expression data were obtained from flybase.

R script.

library(seqinr)

library(Biostrings)

```
>fasta_file <- read.fasta("Soller_UTRs.fa", as.string = T)# read fasta file
```

```
>pattern <- "atg" # the pattern to look for
```

```
>dict <- PDict(pattern, max.mismatch = 0)#make a dictionary of the pattern to look for
```

```
>seq <- DNAStringSet(unlist(fasta_file)[1:638])#make the DNAStringset from the DNAsequences that is, all 638 UTRs related to the 156 genes identified in spanki
```

```
>result <- vcountPDict(dict,seq)#count the pattern in each of the sequences
```

```
>write.csv2(result, "result.csv")
```

```
>fasta_file <- read.fasta("dmel-all-five_prime_UTR-r6.07.fa", as.string = T)# read fasta file
```

```
>pattern <- "atg" # the pattern to look for
```

```
>dict <- PDict(pattern, max.mismatch = 0)#make a dictionary of the pattern to look for
```

```
>seq <- DNAStringSet(unlist(fasta_file)[1:29822])#make the DNAStringset from the DNAsequences that is, all UTRs
```

```
>result <- vcountPDict(dict,seq)#count the pattern in each of the sequences
```

```
>write.csv2(result, "result_allutrs.csv")
```

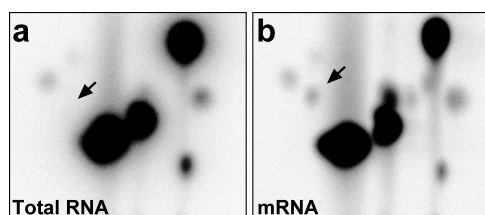
Polytene chromosome preparations and stainings. *Ime4* or *YT521-B* were expressed in salivary glands with *C155-GAL4* from a *UAS* transgene. Larvae were grown at 18°C under non-crowded conditions. Salivary glands were dissected in PBS containing 4% formaldehyde and 1% Triton X-100, and fixed for 5 min, and then for another 2 min in 50% acetic acid containing 4% formaldehyde, before placing them in lactic acid (lactic acid:water:acetic acid, 1:2:3). Chromosomes were then spread under a siliconized cover slip and the cover slip removed after freezing. Chromosomes were blocked in PBT containing 0.2% BSA and 5% goat serum and sequentially incubated with primary antibodies (mouse anti-PolII H5, 1:1000, Abcam, or rabbit anti-histone H4, 1:200, Santa-Cruz, and rat anti-HA monoclonal antibody 3F10, 1:50, Roche) followed by incubation with Alexa488- and/or Alexa647-coupled secondary antibodies (Molecular Probes) including DAPI ($1\mu\text{g ml}^{-1}$, Sigma). RNase A treatment (4 and $200\mu\text{g ml}^{-1}$) was done before fixation for 5 min. Ovaries were analysed as previously described⁴¹.

RNA binding assays. The YTH domain (amino acids 207–423) was PCR-amplified with oligos YTHdom F1 (CAGGGGCCCCCTGTCGACTAGTCCCGGGAA TGGTGGCGGCAACGGCCG) and R1 (CAGCATGAATTGGCGCCGCTCTAGA TTACTTGATAGTACGCTGTATACCTTTTCTCTCGC) and cloned with Gibson assembly (NEB) into a modified pGEX expression vector to express a GST-tagged fusion protein. The YTH domain was cleaved while GST was bound to beads using Precession protease. Electrophoretic mobility shift assays and UV cross-linking assays were performed as described^{35,42}. Quantification was done using ImageQuant (BioRad) by measuring free RNA substrate to calculate bound RNA from input. All binding assays were done at least in triplicates.

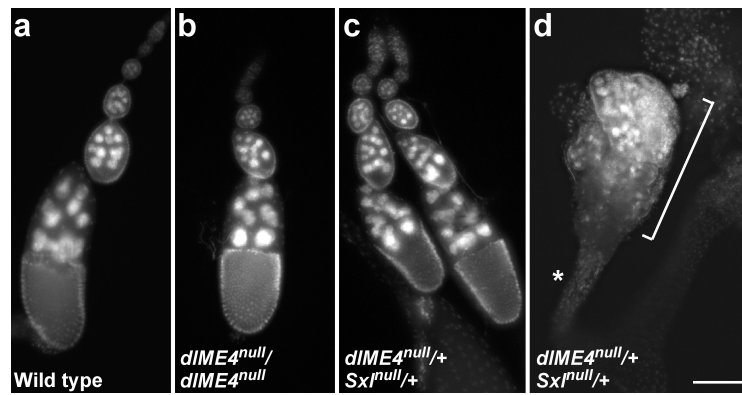
Data availability statement. RNA-seq data that support the findings of this study have been deposited at GEO under the accession number GSE79000, combining the single-end (GSE78999) and paired-end (GSE78992) experiments. All other data generated or analysed during this study are included in this published article and its Supplementary Information.

30. Haussmann, I. U., Hernani, Y., Wijesekera, T., Dauwalder, B. & Soller, M. Multiple pathways mediate the sex-peptide-regulated switch in female *Drosophila* reproductive behaviours. *Proc. R. Soc. Lond. B* **280**, 20131938 (2013).

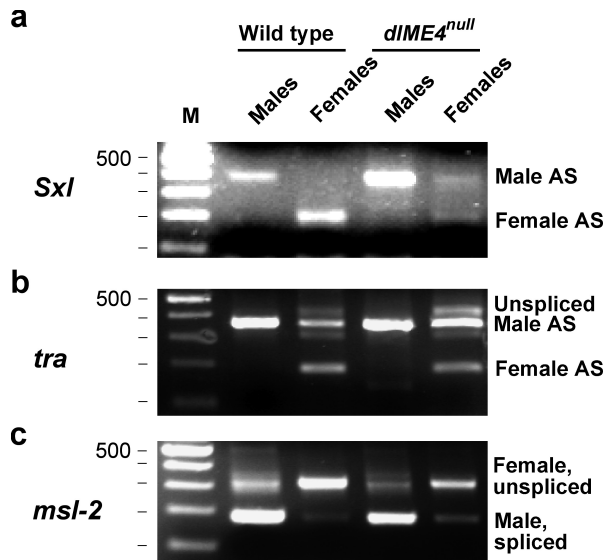
31. Haussmann, I. U., Li, M. & Soller, M. ELAV-mediated 3'-end processing of *ewg* transcripts is evolutionarily conserved despite sequence degeneration of the ELAV-binding site. *Genetics* **189**, 97–107 (2011).
32. Yan, D. & Perrimon, N. *spenito* is required for sex determination in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **112**, 11606–11611 (2015).
33. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25**, 402–408 (2001).
34. Haussmann, I. U., White, K. & Soller, M. Erect wing regulates synaptic growth in *Drosophila* by integration of multiple signaling pathways. *Genome Biol.* **9**, R73 (2008).
35. Soller, M. & White, K. ELAV inhibits 3'-end processing to promote neural splicing of *ewg* pre-mRNA. *Genes Dev.* **17**, 2526–2538 (2003).
36. Harper, J. E., Miceli, S. M., Roberts, R. J. & Manley, J. L. Sequence specificity of the human mRNA N6-adenosine methylase *in vitro*. *Nucleic Acids Res.* **18**, 5735–5741 (1990).
37. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
38. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protocols* **7**, 562–578 (2012).
39. Sturgill, D. *et al.* Design of RNA splicing analysis null models for post hoc filtering of *Drosophila* head RNA-seq data with the splicing analysis kit (Spanki). *BMC Bioinformatics* **14**, 320 (2013).
40. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
41. Soller, M., Bownes, M. & Kubli, E. Control of oocyte maturation in sexually mature *Drosophila* females. *Dev. Biol.* **208**, 337–351 (1999).
42. Soller, M. & White, K. ELAV multimerizes on conserved AU4-6 motifs important for *ewg* splicing regulation. *Mol. Cell. Biol.* **25**, 7580–7591 (2005).



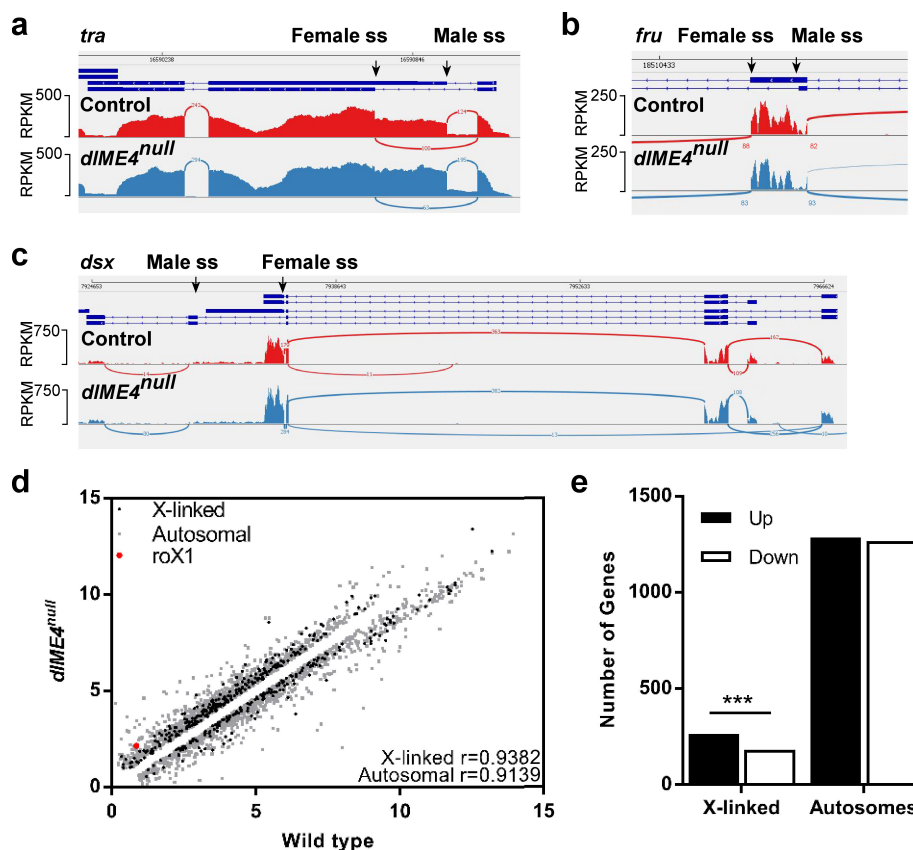
Extended Data Figure 1 | m⁶A levels in unfertilized eggs. a, b, Thin-layer chromatography from maternal total RNA (a) and mRNA (b) present in unfertilized eggs. The arrow indicates m⁶A.



Extended Data Figure 2 | Ime4 supports Sxl in directing germline differentiation. **a–c**, Representative ovarioles of wild-type (**a**), *Ime4^{null}/Ime4^{null}* (**b**) and *Sxl/+;Ime4^{null}/+* females (**c**), and a tumorous ovary of a *Sxl/+;Ime4^{null}/+* female (**d**). The tumorous ovary consisting mostly of undifferentiated germ cells in **d** is indicated with a bracket and the oviduct with an asterisk. Scale bar, 100 μ m (applies to all panels).

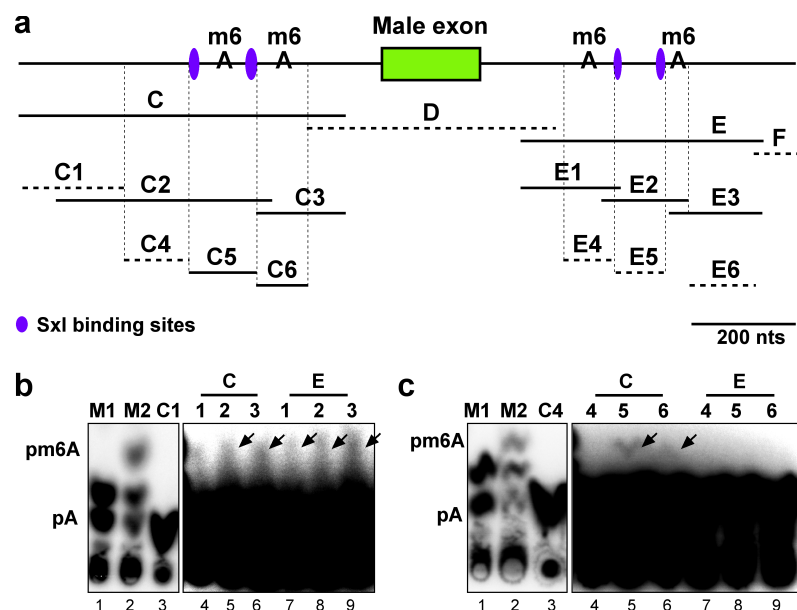


Extended Data Figure 3 | *Ime4* is required for female-specific splicing of *Sxl*, *tra* and *msl-2*. a–c, RT-PCR of *Sxl* (a), *tra* (b) and *msl-2* (c) sex-specific splicing in wild-type males and females, and *Ime4^{null}* males and females. 100-bp markers are shown on the left. AS, alternative splicing.



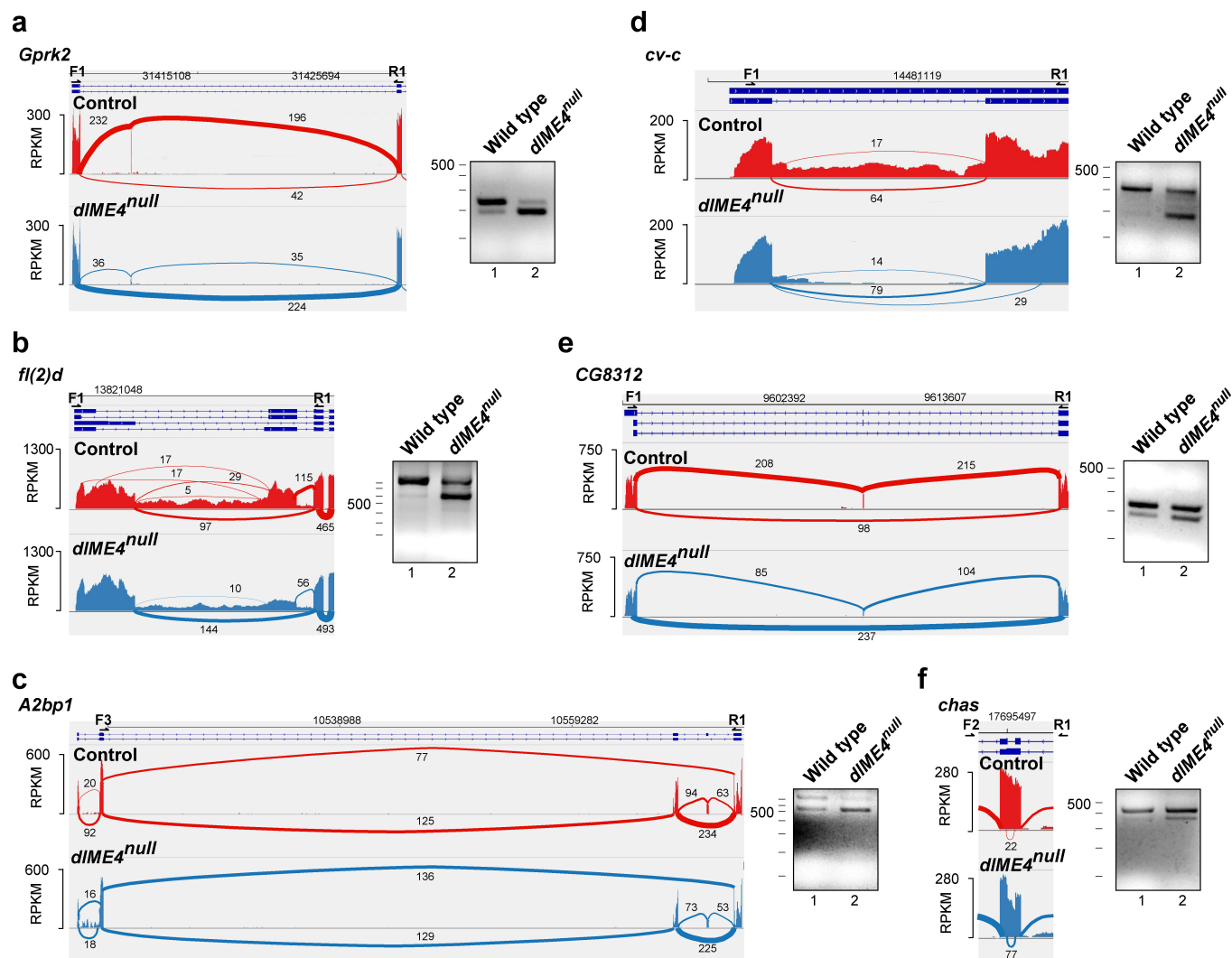
Extended Data Figure 4 | Alternative splicing of sex-determination genes and differential expression of X-linked genes in *Ime4*^{null} females. **a–c**, Sashimi plot depicting Tophat-mapped RNA sequencing reads and exon junction reads below the annotated gene model for sex-specific alternative splicing of *tra*, *fru* and *dsx*. The thickness of lines connecting splice junctions corresponds to the number of junction reads also shown. ss, splice site. **d**, Significantly ($P < 0.05$, $q < 0.166853$) differentially expressed gene expression values expressed as reads per kb of transcript per million mapped reads (RPKM) were $\log[x + 1]$ -transformed and

Spearman r correlation values determined for X-linked and autosomal genes in wild-type and *Ime4*^{null} *Drosophila*. **e**, The proportion of autosomal and X-linked genes that were significantly either up- or downregulated in *Ime4*^{null} as compared to wild-type *Drosophila* were statistically compared using χ^2 with Yates' continuity correction. GraphPad Prism was used for statistical comparisons. Similar results as for the single-read RNA-seq experiment were obtained for the paired-end RNA sequencing experiment.



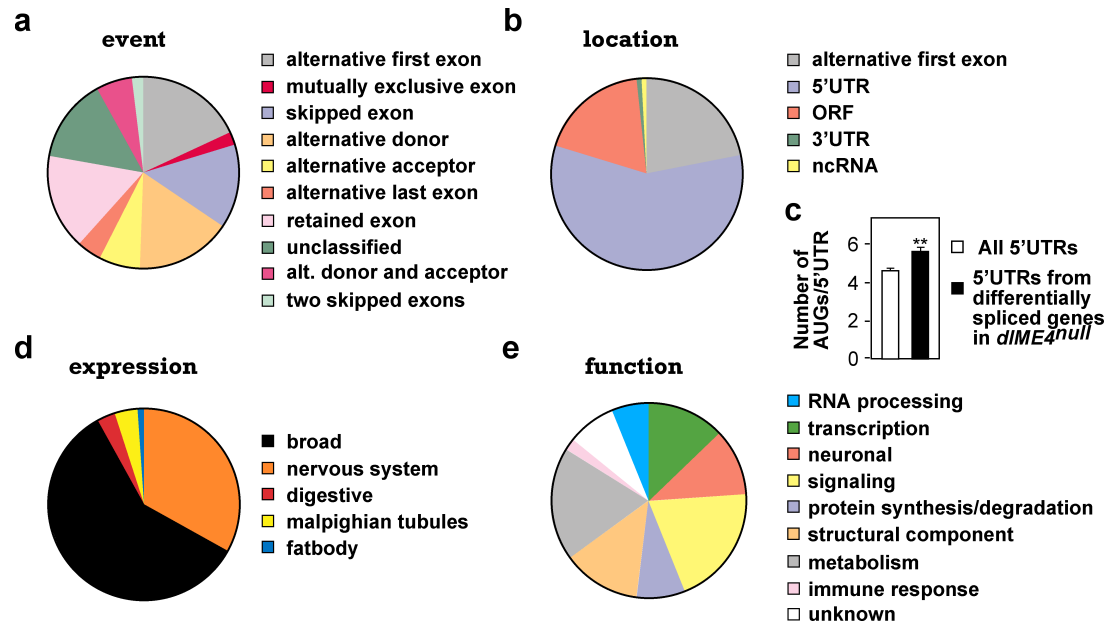
Extended Data Figure 5 | m⁶A methylation sites map to the vicinity of Sxl binding sites. **a**, Schematic of the Sxl alternatively spliced intron around the male-specific exon depicting substrate RNAs used for *in vitro* m⁶A methylation. Solid lines depict fragments containing m⁶A methylation and dashed lines indicate fragments where m⁶A was

absent. **b**, **c**, 1D-TLC of *in vitro* methylated [³²P]-ATP-labelled substrate RNAs shown in **a**. Markers are *in vitro* transcripts in the absence (M1) or presence (M2) of m⁶A ³²P-labelled after RNase T1 digestion. The right panels in **b** and **c** show an overexposure of the same thin-layer chromatography.



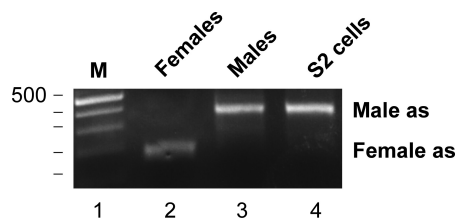
Extended Data Figure 6 | RT-PCR validation of differential alternative splicing in *Ime4*^{null} flies. a–f, Sashimi plots depicting Tophat-mapped RNA sequencing reads and exon junction reads below the annotated gene model of indicated genes on the left, and RT-PCR of alternative splicing

shown on the right using primers depicted on top. The thickness of lines connecting splice junctions corresponds to the number of junction reads also shown.

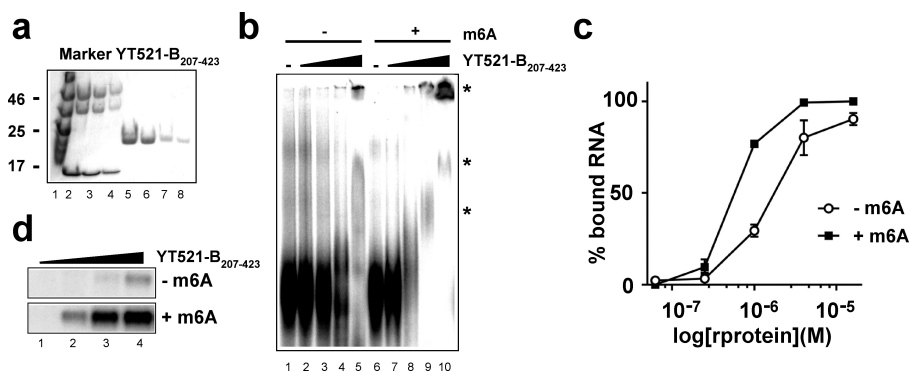


Extended Data Figure 7 | *Ime4* affects alternative splicing predominantly in 5' UTRs in genes with a higher than average number of upstream start codons. **a, b,** Classification of differential alternative splicing in *Ime4^{null}* according to splicing event (**a**) and location of the event in the mRNA (**b**). **c,** Quantification of upstream start codons (AUGs) in all annotated 5' UTRs (white) or in alternative isoforms differentially spliced between wild-type and *Ime4^{null}* insects. All *Drosophila* UTRs were accessed in fasta format from Flybase (version r6.07), (<ftp://ftp.flybase.net/>

[genomes/Drosophila_melanogaster/current/fasta/](http://genomes.Drosophila_melanogaster/current/fasta/)). An R script was used to count the number of ATG sequences in all *Drosophila* 5' UTRs and from the genes identified by the Spanki analysis comprising 638 5' UTRs. A *t*-test was then used to statistically compare the number of ATGs present in the 638 5' UTRs of the differentially spliced genes as compared to all 29,822 *Drosophila* 5' UTRs. **d, e,** Classification of differentially alternatively spliced genes in *Ime4^{null}* according to expression pattern (**d**) or function (**e**).

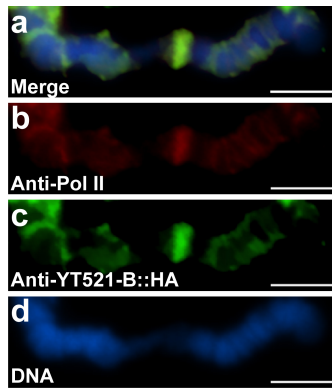


Extended Data Figure 8 | *Drosophila* S2 cells are male. RT-PCR of *Sxl* alternative splicing in females, males and S2 cells. 100-bp markers are shown on the left.



Extended Data Figure 9 | Preferential binding of the YTH domain of YT521-B to m⁶A-containing RNA. **a**, Coomassie-stained gel depicting the recombinant YTH domain (amino acids 207–423) of YT521-B. **b**, **c**, Electrophoretic mobility shift assay of YTH domain binding to *Sxl* RNA fragment C with or without m⁶A (50% of adenosine in the transcript methylated) and quantification of RNA bound to the YTH domain shown

as mean \pm s.e.m. ($n = 3$). Note that the YTH domain does not form a stable complex with RNA (asterisk) and that this complex falls apart during the run or forms aggregates in the well. **d**, UV cross-linking of the YTH domain to *Sxl* RNA fragment C at 0.25 μ M, 1 μ M, 4 μ M and 16 μ M (lanes 1–4).



Extended Data Figure 10 | YT521-B co-localizes to sites of transcription. **a–d**, Polytene chromosomes from salivary glands expressing YT521-B::HA stained with anti-Pol II (red, **b**), anti-HA (green, **c**) and DNA (DAPI, blue, **d**), or merged (yellow, **a**). Scale bars, 5 μ m.

CAREERS

JOB SECURITY How having children can change your priorities **p.306**

BLOG Stories and more to help you reach career goals blogs.nature.com/naturejobs

NATUREJOBS For the latest career listings and advice www.naturejobs.com



ADAPTED FROM GETTY

BY PHILIP CAMPBELL

“Her entire mentoring style is a gift to science.” So said a nominator of Julie Overbaugh, an HIV researcher at the Fred Hutchinson Cancer Research Center in Seattle, Washington. And the judges of this year’s annual *Nature* Awards for Mentoring in Science must have felt the same: they awarded Overbaugh the lifetime-achievement prize. They also awarded the prize for mid-career achievement, to biologist Susan Forsburg at the University of Southern California, Los Angeles.

Founded in 2005, the awards focus on a different region or country each year (see go.nature.com/2eejupo). In 2016, it is the US states of Washington, Oregon and California (see go.nature.com/2jwik45). Each contender has to be nominated by five of their former trainees, covering different stages of the mentor’s career.

An important quality of the award winners is their ability to help protégés to reach crucial milestones. This skill, nominators say, is exemplified by Overbaugh.

“Given her success in the high-pressure environment of retroviral pathogenesis, it is remarkable that her lab has remained an outstanding training environment for graduate students,” says one of her nominators. “She possesses the rare ability to work with her students to identify and develop projects that are relatively ‘protected’ from outside competition, yet retain enough scientific rigour to foster their development. It is hard to find a student of hers who has not published at least one significant contribution to the field.”

Overbaugh shares with previous award winners the capacity to tune her mentoring to the protégé’s individual personality and needs. “Julie has high expectations of her trainees and aims for all of them to develop the fundamental skills of a successful scientist,” says the nominator, adding that Overbaugh does not expect trainees to follow her path into academia: many of them now have important roles in government, philanthropy and industry. “She recognizes each individual’s strengths, weaknesses and goals, and figures out how to help them get to where they want to be,” the nominator says. “The breadth and diversity of roles that her mentees currently fill is a testament to this style of mentorship that is rare even among successful academic mentors.”

Another of Overbaugh’s nominators ►

MENTORING AWARDS

Leading by example

Nature announces this year’s outstanding science mentors from the US west coast.

TRADE TALK

Drug referee



Steffen Schulz was completing his PhD in medical neuroscience at the Charité university hospital in Berlin when he realized he wanted more job security than academia could offer. He is now a

drug-safety manager at a pharmaceutical company in Germany.

Where do you work?

I'm based in Berlin. I grew up, studied for my diploma and did my PhD here, and I'm still in the same area.

How did you get your job?

I actually got it before I finished my PhD. My wife was pregnant with twins and I didn't want another fixed-term contract or to be forced to move anywhere. I considered a postdoc, but that felt like a stopgap. So I applied for work at a large pharmaceutical company, and started a trainee position there before I submitted my thesis. Since then I've been made permanent, and then promoted.

What do you do there?

In the drug-safety department, we take messages from physicians, pharmacists and patients about side effects from our company's drugs. I record and analyse that feedback, evaluate the benefit-to-risk ratio and see if any changes should be made to the patient-information leaflet or if we need to do something more dramatic.

Such as?

The serious side effects are reported directly to health agencies — Germany's Federal Institute for Drugs and Medical Devices and the European Medical Agency. They can withdraw a drug from the market.

What would you tell people who want to move out of academia?

You can't apply early enough — it's worth trying even before you submit your thesis. Be aware of the different options. I hadn't heard of drug safety before I saw this job. And know your skills — most PhD students don't realize that they already have a lot of experience. Even managing experiments is a form of project management. That's valuable. ■

INTERVIEW BY JACK LEEMING

This interview has been edited for length and clarity. For more, see go.nature.com/2fv6p2s

► speaks highly of her engagement with students from other countries, especially through her collaborations in projects based in Kenya. "She has an excellent capability and patience in guiding international scholars," this nominator says. "She has the patience to listen to and deal with culture shocks and adjustment to new surroundings and a different system of training and education."

Excellent mentoring can take different forms, as Forsburg's trainees have found. Imagine you're a graduate student who is frustrated by your lack of progress in understanding a protein kinase. Meanwhile, a talented postdoc in your lab characterizes a new role for it that leads to a strong paper. You're envious, embarrassed that none of your own work is included in the paper and express your frustration to your supervisor. What happens next?

A student mentored by Forsburg faced precisely this issue. What was Forsburg's advice? "Susan took a firm and constructive stance with me — challenging me to be more deliberate in my work, and to use my postdoc colleague as a role model," says the student, who was one of Forsburg's nominators. "She didn't give me any easy answers, but instead gave me a role model and higher expectations. At the same time, she also gave me more time and attention to help me think through my work, and encouraged the postdoc to mentor me to better set me up for success."

This is one of many examples of how Forsburg rises to the challenge of nurturing the career progress of younger scientists. She has gained a scientific reputation in pioneering the use of single-cell analysis and live-cell imaging by investigating mechanisms that maintain genome stability during stresses in replication, using fission yeast as a model system. But she was recognized for her strong role in helping undergraduates, graduate students and postdocs to fulfil their scientific potential.

Forsburg was a stand-out nominee, according to Lisa Coussens, chair of the judging panel and of the cell, developmental and cancer biology department at Oregon Health & Science University in Portland. She says that Forsburg not only strongly mentors members of her own laboratory, as evidenced by anecdotes and testimonials from her nominators, but also engages with mentoring challenges on a grander scale.

As a prominent advocate for women in science, Forsburg maintains a career-oriented website, Women in Biology (www.womenbio.net), which provides information on career management for female scientists and young science professionals. She has also written many columns for the 'Women in Cell Biology' feature of the American Society for Cell Biology newsletter.

Another of Forsburg's nominators says that the biologist treats junior researchers in her lab as much more than available labour.



ROBERT HOOD/FRED HUTCHINSON CANCER RESEARCH CENTER



S. L. FORSBURG

Julie Overbaugh (top) and Susan Forsburg won this year's *Nature* mentoring awards.

"She helps students take a formal, logical approach in trying to understand problems and develop options to solve them, having students develop robust hypotheses and likely outcomes of experiments before performing them," the nominator says. "This approach requires students to take intellectual ownership of their work and creates students that are trained to think as opposed to 'skilled pairs of hands'."

The nominator also notes that Forsburg encourages her protégés to broaden their scientific perspective. "Susan's expectation is that lab members should have a working understanding and ability to contribute across all areas," the nominator says. "This will give her lab members greater ability to understand and pursue surprising or novel results."

Outstanding mentors share many qualities (see *Nature* 447, 791–797; 2007). To judge from their nominations, both of this year's winners of *Nature*'s mentoring awards are no exception. ■

Philip Campbell is editor-in-chief of *Nature*.

between two voices talking

Brief encounter.

BY LAURENCE RAPHAEL BROTHERS

The alien vessel entered orbit around the Moon and maintained silence, ignoring all transmissions. The eyes of the world were fixed on the craft as a joint Russian–American mission set out to make contact.

With the voyage capsule approaching its rendezvous, the two crew members prepared for their historic encounter. NASA's Marla Pham, a specialist in languages, was charged with boarding the alien vessel. Colonel Pyotr Ulyanovich Simonov of Roscosmos remained in the capsule.

Marla hadn't mentioned anxiety attacks to the counsellor during her evaluation. The last had been years ago, before her orals. But this attack had been building ever since launch. And now, at the crucial moment, it crested. Fear filled Marla's being, rushed through her blood vessels, pulsed through her nerves. It was almost paralysing, but she forced herself to pass through the docking tube and into the unknown.

The alien's outer airlock door swung shut as Marla entered; she couldn't help shuddering as her path to safety was cut off. The tiny cramped chamber cycled to fill with air, and the inner door opened to reveal a surprisingly homely little room.

Artificial gravity: 1.0 G. Temperature: a balmy 26°C. Appearance: dark-blue carpeting with an ornate yellow geometrical pattern. Equipment: a low wooden table in the centre of the room; a black ceramic teapot on the table, steam rising from the spout. Two off-white porcelain cups sat by the pot, hexagon patterns in faint cobalt blue glazed into the finish. Illumination came from the ceiling, which Mission Control said duplicated Sol's spectrum.

Built for humans, Marla thought.

The room was unshielded; her comm links stayed up. Overcoming a terrible moment of total paralysis, Marla cracked the seal on her helmet.

"Marla! What are you doing? Are you crazy?" Pyotr's voice buzzed in her ear.

I'm in charge of contact, she sent, texting him from a tablet on her forearm to avoid speaking out loud.

NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

Despite Pyotr's protests, despite her own fear, Marla removed her spacesuit and laid

it by the airlock. She couldn't drink tea in armour, no matter how high the risk. Grabbing her tablet, Marla walked cautiously to the table and knelt down. She sniffed the air and was rewarded with the fragrant scent of *trà sen* lotus tea.

Vietnamese tea! Could this have been prepared just for me?



"Thank you for arranging this room," she said in English, a small tremor in her voice. "It is very comfortable."

Silence. It felt like hours but her lens display said less than a minute had passed. Suggestions from Control scrolled by in the lower field of her vision. They could tell her heart rate was 120, but they thought it was normal nerves.

"Nothing happening," said Pyotr. "There are no signals."

At last she said in Mandarin: "I welcome you to our world."

No response.

"I greet you in friendship," she said, switching to French.

She paused.

"Nothing," said Pyotr. "You should come back."

She wanted desperately to agree but knew she'd never forgive herself if she did.

"Please answer," she said in Spanish. "You have made the long journey to our world. You must have created this room for a reason."

Still no response.

"Maybe it's just a machine," Pyotr suggested.

Yes... perhaps this whole spacecraft is nothing more than a trap. Marla swallowed, forcing the notion away.

She switched to Vietnamese. "You must have wanted this meeting. Please —" Her voice caught in her throat.

Nothing but silence.

"Maybe it's shy," said Pyotr.

Marla bit her lip.

You might be right, she keyed on her tablet.

"Ha," he said, "you think?"

She stood up and bowed.

"I'm sorry," she said. "I came to meet you. But I feel fear as well. Fear of the unknown. I understand if you hesitate to show yourself.

Perhaps my appearance scares you. But it seems a shame to come all this way for nothing."

She bowed again, gathered up her suit and turned towards the airlock. Was there even a way to open the door?

"Bozhe moi! Turn round, Marla!"

Marla's fear surged up again like a geyser, blasting through her like fire and ice and shame and death. It was the hardest thing she'd ever done not to scream. Because now she knew she was not alone.

The voice came from behind her. Spoken in Vietnamese, so softly she could barely hear it.

"Wait," it whispered. "Please..."

Marla took a deep breath and turned round.

There was a shadow on the far side of the room. It was small and she couldn't see what it concealed. Marla put down the spacesuit and returned to kneel before the table, her heart hammering in her chest. She extended her hand slowly, placing it palm up on the table, feeling the blood racing, making her fingertips shudder with every beat. The shadow extended towards her. She felt a feathery delicate touch like a young fern frond tentatively exploring her hand. The shadow encompassed the teapot, and when it withdrew, two cups had been poured.

"Let us speak together," said Marla.

"I was afraid," said the voice. "I still am."

"So was I," she said. "So am I still."

"It was an unworthy fear."

"Yes."

"You felt it too?"

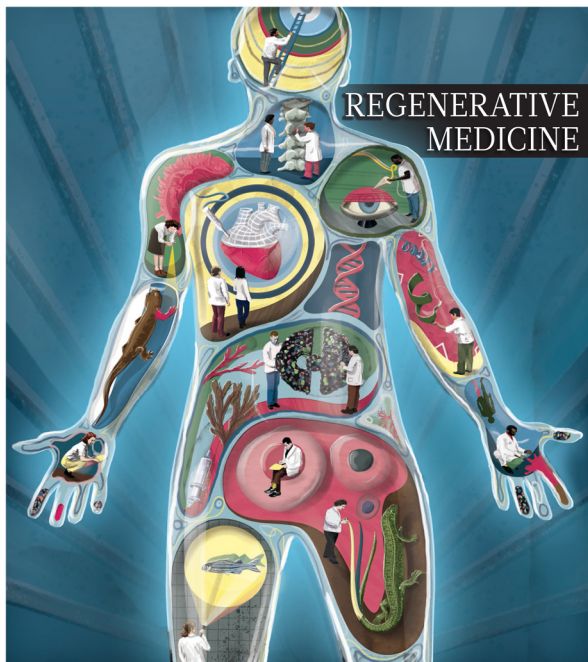
"I did."

"Are we the same?"

"Close enough," she said, and she picked up her cup, the warm welcoming scent curling up around her. Across the table, the shadow faded away; but her host remained. Together, they drank their tea. ■

Laurence Raphael Brothers is a writer and a technologist. His stories have appeared in such venues as *Daily Science Fiction*, *The Sockdolager* and *The New Haven Review*. Follow him on Twitter: @lbrothers

ILLUSTRATION BY JACEY



Produced with support from:



Rebuilding
the body

natureOUTLOOK

REGENERATIVE MEDICINE

8 December 2016 / Vol 540 / Issue No 7632



Cover art: Sam Falconer

Editorial

Herb Brody, Michelle Grayson, Richard Hodson, Jenny Rooke

Art & Design

Mohamed Ashour, Kate Duncan, Wesley Fernandes, Wojtek Urbanek

Production

Matthew Carey, Ian Pope, Karl Smart

Sponsorship

Yuki Fujiwara, Yvette Smith

Marketing

Nicole Jackson

Project Manager

Anastasia Panoutsou

Art Director

Kelly Buckheit Krause

Publisher

Richard Hughes

Editorial Director,

Partnership Media

Stephen Pincock

Chief Magazine Editor

Helen Pearson

Editor-in-Chief

Philip Campbell

Bodies wear out. Tissue thins and tears. Organs stop functioning. Cells lose their biological way. Trauma breaks things. And as a result, we become ill or disabled. This has always been our fate.

Regenerative medicine is the bold collection of techniques and technologies that aim to restore our physiology to something that resembles its original condition. Its roots trace back to antiquity (see page S50) but it has, in recent years, become much more effective. For example, 3D printers can construct tissue and organs that in some cases can function as well as the originals (S56). The central nervous system, however, has proved stubbornly difficult to repair. Scientists hope that stem-cell advances might finally restore mobility to those with spinal-cord injuries (S52). Excitement is already building over the potential of these intriguing cells to create drug-free treatments for chronic diseases such as type 1 diabetes (S60). Regeneration researchers are also taking cues from the animal world: species such as salamanders have the power to regrow limbs. Understanding the cellular mechanisms behind this ability might lead to techniques that can work in humans (S58). Indeed, one scientist argues that for progress to continue, researchers will need to do a better job of emulating and working alongside natural systems (S55).

The creation of these therapies brings with it questions of how to regulate them. Clinics are sprouting up to offer dubious stem-cell-based treatments for dire conditions, and policymakers are crafting rules to accelerate the availability of effective treatments without endangering desperate patients (S64).

We are pleased to acknowledge financial support for this Outlook from Translational Research Informatics Center (TRI), Clio, Inc., Sapporo Medical University and CYBERDYNE, INC. As always, *Nature* retains sole responsibility for all editorial content.

Herb Brody

Chief supplements editor

CONTENTS

S50 TIMELINE

Regrowing the body

The big breakthroughs in our understanding of regeneration

S52 NEUROSCIENCE

New nerves for old

Stem-cell treatments have resulted in some surprising recoveries, but neurologists are cautious

S55 PERSPECTIVE

Work with, not against, biology

Tissue engineers need to rethink their approach, says Stephen Badylak

S56 TECHNOLOGY

The promise of printing

Perfecting the techniques that could bring printed organs to the clinic

S58 ANIMAL MODELS

Unlock your inner salamander

Learning from nature's regenerators

S60 DIABETES

Encapsulating the problem

Developing devices to shield implanted cells from the immune system

S64 REGULATION

Rewriting the regenerative rulebook

The controversy surrounding regulation of stem-cell therapy

RELATED ARTICLES

S68 An FDA perspective on preclinical development of cell-based regenerative medicine

Alexander M. Bailey, Michael Mendicino & Patrick Au

S71 Autophagy maintains stemness by preventing senescence

Laura García-Prat et al.

S77 Modulation of tissue repair by regeneration enhancer elements

Junsu Kang et al.

S83 Spinal cord reconstitution with homologous neural grafts enables robust corticospinal regeneration

Ken Kadoya et al.

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the *Nature Outlook* Editorial guidelines available at go.nature.com/e4dwzw

CITING THE OUTLOOK

Cite as a supplement to *Nature*, for example, *Nature* Vol. XXX, No. XXXX Suppl., Sxx–Sxx (2016).

VISIT THE OUTLOOK ONLINE

The *Nature Outlook Regenerative Medicine* supplement can be found at <http://www.nature.com/nature/outlook/regenerative-medicine>. It features all newly commissioned content as well as a selection of relevant previously published material.

All featured articles will be freely available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES

Site licences (www.nature.com/libraries/site_licences): Americas, institutions@natureny.com; Asia-Pacific, <http://nature.asia/jp-contact>; Australia/New Zealand, nature@macmillan.com.au; Europe/ROW, institutions@nature.com; India, npgindia@nature.com. Personal subscriptions: UK/Europe/ROW, subscriptions@nature.com; USA/Canada/Latin America, subscriptions@us.nature.com; Japan, <http://nature.asia/jp-contact>; China, <http://nature.asia/china-subscribe>; Korea, www.natureasia.com/ko-kr/subscribe.

CUSTOMER SERVICES

Feedback@nature.com

Copyright © 2016 Macmillan Publishers Ltd. All rights reserved.

Regrowing the body



◀◀◀ ~600 BC

In one of the first surgical textbooks, Indian surgeon Suśruta (**pictured**) discusses skin-graft techniques. The book details a method for repairing torn earlobes with skin from the cheek and includes the first written record of reconstruction of the nose from a flap of forehead skin.



1952

US scientists Robert Briggs and Thomas King report the first cloning using nuclear transfer. The researchers remove the nucleus from a frog egg and replace it with the nucleus from a frog-embryo cell. The egg develops

into a tadpole. The experiment shows that nuclei hold onto the organism's genome even after they have been transferred into a new cell body. Briggs's initial request for funds is dismissed by the US National Cancer Institute, which called nuclear transplantation a "hare-brained scheme".

1981

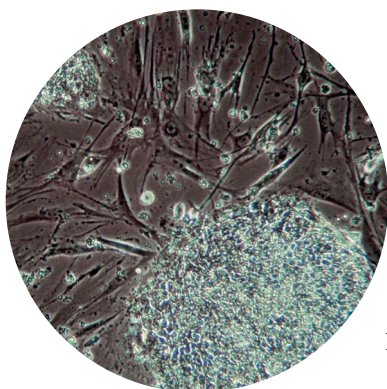
UK researchers Martin Evans and Matthew Kaufman are the first to isolate stem cells from mouse embryos. In the same year, US scientist Gail Martin discovers how to keep stem cells alive in a petri dish, enabling experimentation.

1997

A paper by Harvard Medical School researcher Joseph Vacanti and his team includes an image of what appears to be a human ear growing out of a mouse's back. The team hopes that its research will lead to lab-grown ears for people whose ears are damaged. But the image sparks a backlash from animal rights activists and people who find the image shocking.

1998

Two teams announce that they have isolated human embryonic stem cells (**pictured**). James Thomson at the University of Wisconsin–Madison and his colleagues derive cells from donated embryos originally intended for *in vitro* fertilization. And a team led by John Gearhart from Johns Hopkins University in Baltimore, Maryland, uses embryos from terminated pregnancies.

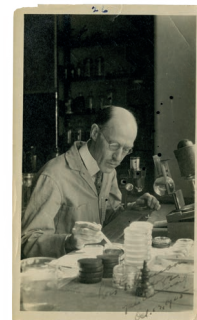


1740 AQUATIC DISCOVERY

Swiss naturalist Abraham Trembley, 'the father of biology', discovers that freshwater hydra can regenerate.

1901 HEREDITY MADE CLEAR

Thomas Hunt Morgan, a US scientist best known for his fruit-fly research on the role of chromosomes in inheritance, writes *Regeneration*. The book helps to bring order and clarity to a field rife with inconsistencies and confusing terminology.



1907 ENGINEERING ORIGINS

US biologist Ross Granville Harrison (**pictured**) discovers a technique for growing embryonic frog cells in the lab. His experiment, the first stem-cell experiment and the first successful method for culturing tissue, paves the way for tissue engineering.

1963 STEM-CELL PROOF

Canadian scientists James Till and Ernest McCulloch identify stem cells in mouse bone marrow that can self-renew and mature into platelets and red and white blood cells.

1981 REPLACEMENT SKIN

Biologist Eugene Bell and his colleagues report a method for repairing wounds with artificial skin made from an individual's own cells. Bell then founds Organogenesis, which in 1998 becomes the first company to receive US Food and Drug Administration approval for a medical product containing living cells.

CLOCKWISE FROM TOP LEFT: PARIS PIERCE/ALAMY; JOEL SARTORE/GETTY; EMBRYO PROJECT/MBL ARCHIVES/ARIZONA BOARD OF REGENTS; JEFF MILLER

For centuries, scientists have been captivated by the phenomenal feats of regeneration found in nature. Despite decades of research, attempts to replace or repair parts of the human body have met with only modest success. Fresh understanding of organ formation coupled with new technologies may help to unlock long-sought cures. By Cassandra Willyard

2001 OPPOSITE ACTS

US President George W. Bush prohibits federal funds being used to start new human embryonic cell lines or to conduct research on newly derived cell lines. The same year, the United Kingdom loosens restrictions on embryo research.

2006 BUILDING BODY PARTS

Anthony Atala, a surgeon at Wake Forest School of Medicine in Winston-Salem, North Carolina, successfully transplants lab-grown bladders into seven children and teenagers with congenital birth defects.

2009 ORDER OVERTURNED

US President Barack Obama lifts the ban on federal funding of new human embryonic stem cell lines.

2010 LARGE-SCALE SUCCESS

An individual with a spinal-cord injury becomes the first person to receive a therapy derived from embryonic stem cells. A large study the same year reports that stem cells harvested from patients' healthy cornea tissue can be grown in the lab and transplanted into the damaged corneas to restore eyesight.

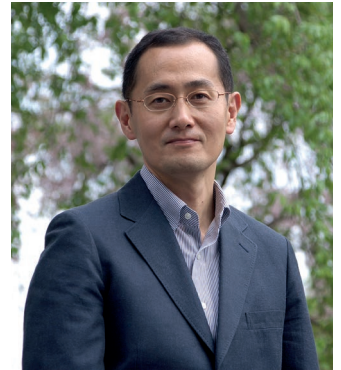


2015 FIRST TO MARKET

The Europe Commission approves the sale of Holoclar to treat people with severely damaged corneas. It is the first stem-cell therapy to reach the market.

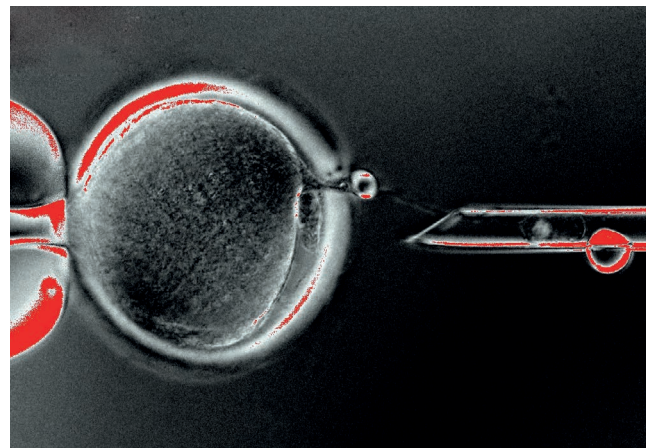
2006

Shinya Yamanaka (**pictured**) from Kyoto University in Japan coaxes adult mouse cells to return to an embryonic-like state by adding four transcription factors. These induced pluripotent stem (iPS) cells circumvent the ethical issues that surround embryonic stem cells. The year after, teams led by Yamanaka and James Thomson manage to reprogram adult human cells, and a group that includes more than 100 British medical charities writes a letter to *The Times* newspaper warning patients to be wary of stem-cell wonder cures.



2013

Shoukhrat Mitalipov at Oregon Health and Science University in Portland and his colleagues create the first human stem-cell lines by therapeutic cloning. The process involves somatic-cell nuclear transfer (**pictured**) — the same technique used to clone Dolly the sheep. First, researchers take the nucleus from a donor cell and inject it into an egg from which the nucleus has been removed. The researchers then coax the egg into dividing, harvest stem cells from the resulting embryo, and then culture those cells. Mitalipov's team used donor cells from a baby. The following year, two research



teams — one based in the United States and one in South Korea — report that they created human embryonic stem cells by cloning adult cells. This kind of therapeutic cloning could help researchers to develop genetically matched stem-cell therapies that are less likely to be rejected.

2014

A team of US and Japanese researchers reports that it can coax cells into an embryonic state by bathing the cells in acid, a technique called stimulus-triggered acquisition of pluripotency or STAP. The discovery promises an easy way to create patient-specific stem cells. But later that year, *Nature* retracts the two papers after it becomes clear that some of the data were falsified.

2016

Japanese scientists led by Katsuhiko Hayashi of Kyushu University in Fukuoka transform mouse skin cells into eggs by culturing them in clumps of fetal mouse ovary tissue. When fertilized and transplanted into a surrogate, these lab-grown eggs give rise to healthy, fertile pups.



Kristopher Boesen, who broke his neck in a car accident, regained the ability to move his arms and hands after his spinal cord was injected with stem cells.

NEUROSCIENCE

New nerves for old

Stem-cell therapy promises to restore motor function after a stroke or spinal-cord injury, but neurologists are proceeding with caution.

BY KATHERINE BOURZAC

Two years after having a stroke at 31, Sonia Olea Coontz remained partially paralysed on her right side. She could barely move her arm, had slurred speech and needed a wheelchair to get around. In 2013, Coontz enrolled in a small clinical trial. The day after a doctor injected stem cells around the site of her stroke, she was able to lift her arm up over her head and speak clearly. Now she no longer uses a wheelchair and, at 36, is pregnant with her first child.

Coontz is one of stem-cell therapy's "miracle patients", says Gary Steinberg, chair of neurosurgery at Stanford School of Medicine in California, and Coontz's doctor. Conventional wisdom said that her response was impossible: the neural circuits damaged by the stroke were dead. Most neuroscientists believed that

the window for functional recovery extends to only six months after the injury.

Stem-cell therapies have shown great promise in the repair of brain and spinal injuries in animals. But animal models often behave differently from humans — nervous-system injuries in rats, for example, heal more readily than they do in people. Clinical trial results have been mixed. Interesting signals from small trials have faded away in larger ones. There are plenty of unknowns: which stem cells are the right ones to use, what the cells are doing when they work and how soon after an injury they can be used.

The field is still young. Stem cells are poorly understood, and so is what happens after a spinal-cord injury or stroke. Yet, there are success stories, such as Coontz's, which seem to show that therapy using the right sort of stem cell can lead to functional improvements when

tried in the right patients and at the right time following an injury. Researchers are fired up to determine whether stem-cell therapies can help people who are paralysed to regain some speech and motor control — and if so, what exactly is going on.

SCAR TISSUE

Neurologists seeking functional restoration are up against the limited ability of the human central nervous system to heal. The biology of the brain and spinal cord seems to work against neuroregeneration, possibly because overgrowth of nerves could lead to faulty connections in the finely patterned architecture of the brain and spine, says Mark Tuszynski, a neurologist at the University of California, San Diego. Local chemical signals in the central nervous system tamp down growth. Over time, scarring develops, which prevents the

GREG IGER/KECK MEDICINE OF USC

injury from spreading, but also keeps cells from entering the site.

“It’s really hard to fix the biology,” says Charles Yu Liu, a neurosurgeon and director of the University of Southern California Neurorestoration Center in Los Angeles. Stem cells seem to promise a workaround.

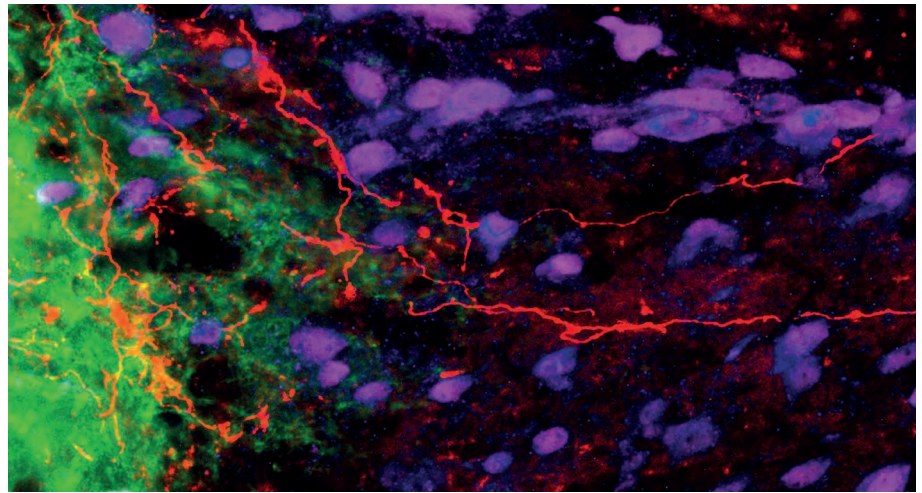
So far, neural regeneration cell therapy has had only anecdotal success, leaving investors and patients disappointed. In people with Parkinson’s disease, for example, neurosurgeons replaced dead and dying dopamine-producing neurons with fetal neurons. Although initial results were promising, in larger studies, patients reported involuntary movements. Another effort tried treating people who’d had a stroke with cells derived from tumours; the results were mixed, and researchers were uneasy about the cells’ cancerous source.

In recent years, researchers have had success with stem cells coaxed to develop into particular cell types, such as neural support cells. Tuszynski has showed how well stem cells can work — at least, in animal models¹. His group implanted neural stem cells derived from human fetal tissue into rats with severe spinal-cord injuries. Seven weeks later, the cells had bridged the gap where the spinal cord had been cut and the animals were able to walk again. The cells used in the study were manufactured by Neuralstem of Rockville, Maryland. The group has shown that other kinds of stem cell, including those derived from adult tissue, also work. Tuszynski has seen similar results in a rat spinal-cord-injury model, using neural stem cells made from the tissues of a healthy 86-year-old volunteer².

But animal studies are also making it clear that simply regrowing the connective wiring of the nervous system to bridge damaged areas is not enough, says Zhigang He, who studies neural repair at the Harvard Stem Cell Institute in Cambridge, Massachusetts. No matter what the animal model is, he says, the axons don’t always grow into the right places. It’s not enough to have a nerve, that nerve must become part of a functional circuit.

There is growing evidence that besides becoming replacement nerves, stem cells perform other functions — they also seem to generate a supportive milieu that may encourage the natural recovery process or prevent further damage after an injury. Many types of neural stem cell secrete a mix of molecules that unlock suppressed growth pathways in nerves. Earlier this year, Tuszynski reported that any sort of spinal-cord stem cell, whether derived from adult tissues or embryos, from humans, rats or mice, could trigger native neural regeneration in rats³.

“Every time we get an experiment done we realize it’s more complex than we thought it would be.”



Regeneration of axons (red) beyond implanted neural progenitor cells (green) in a rat with a spinal injury.

But his success in rats has not yet translated into clinical trials. More work is needed, Tuszynski says, to determine which type of cell will work best for which particular injury.

REINVENTING RECOVERY

For people who have had a stroke or spinal-cord injury, physical therapy is currently the best hope for recovery in the weeks and months after the injury. The brain is plastic and can co-opt other circuits and pathways to compensate for damage and to restore function. Once the inflammation ebbs and the brain adjusts, people can start to regain function. But the window of opportunity is short. Most people don’t make functional gains after six months.

That timeline is why the remarkable recovery enjoyed by Coontz and other patients with chronic stroke in the same clinical trial is so surprising, says Steinberg. “This changes our whole notion of recovery,” he says. There were 18 people in the trial Coontz took part in, and all were treated using stem cells manufactured by SanBio of Mountain View, California. The company’s cells are bone-marrow-derived mesenchymal stem cells. The cells are treated with a DNA fragment that is transiently expressed in them, and causes changes in their protein-expression patterns. In animal studies, these cells promote the migration and growth of native neural stem cells, among other effects.

The trial, which was designed to look at safety as well as efficacy, recruited patients after an ischaemic stroke. During this kind of stroke, a clot cuts off the blood supply to part of the brain, causing significant damage. Patients in the trial had all had ischaemic strokes deep in the brain 7–36 months earlier — past the 6-month window for significant recovery. Each patient was injected with either 2.5 million, 5 million or 10 million of SanBio’s cells⁴. Steinberg has followed participants for 24 months; an interim study at 12 months reported that most patients

showed functional improvements. Some, like Coontz, achieved almost complete recovery.

What is not clear, however, is what the stem-cell injections do in the brain. In animal studies, the SanBio cells do not turn into neurons, but seem to send supporting signals to native cells in the brain. Indeed, preclinical research shows that the cells do not integrate into the brain — most die after 12 months. Instead, the cells seem to secrete growth factors that encourage the formation of new neurons and blood vessels, and foster connections called synapses between neurons. And in rats, the nerve-cell connections that extended from one side of the brain to the other, as well as into the spinal cord, lasted, even though the injected cells did not⁴.

But these mechanisms are not sufficient to explain Coontz’s overnight restoration of function, says Steinberg. He is entertaining several hypotheses, including that the needle used to deliver the cells may have had some effect. “One week after treatment, we saw abnormalities in the premotor cortex that went away after one month,” he says. The size of these microlesions was strongly correlated with recovery at 12 months. A similar effect can happen when electrodes are implanted in the brains of people with Parkinson’s, although this deep-brain stimulation quiets tremors for only a short time. The people who’d had a stroke had a lasting recovery, suggesting that both the needle and the stem cells may have played a part.

The SanBio trial was small, and did not have a placebo control; the company is now recruiting for a larger phase II trial. Of the 156 participants that will be recruited, two-thirds will have cells injected — the others will have a sham surgery. Even the trial surgeons, including Steinberg, will not know who is getting which treatment. The main outcome measure will be whether patients’ motor-skill scores improve on a test called the Fugl-Meyer Motor scale six months after treatment. Participants will be monitored for

at least 12 months, and will also be evaluated with tests that look for changes in gait and dexterity. Meanwhile, Steinberg plans to study microlesions in animal models of stroke to determine whether they do have a role in recovery.

SMALL SUCCESSES

An ongoing clinical trial evaluating escalating doses of neural stem cells in patients with acute spinal-cord injuries is also looking promising. Asterias Biotherapeutics of Fremont, California, coaxes the cells to develop into progenitors of oligodendrocytes, a type of support cell that's found in the brain and spinal cord and that creates a protective insulation for neuronal axons.

The trial tests the safety and efficacy of administering these cells to people with recent cervical, or neck-level, spinal-cord injury. Interim results for patients who had received the two lower doses were presented at the International Spinal Cord Society meeting in September. After 90 days, 4 patients who received 10 million cells showed improved motor function; a fifth patient had not reached the 90-day mark yet. At one year, the three patients receiving a lower dose of two million cells showed measurable improvement in motor skills.

These cells were initially developed by Geron, a biotechnology company that has since moved away from regenerative medicine. Before spinning out Asterias in 2013, Geron had run a safety trial of the cells in people with a chronic lower-back injury. No issues were identified, and the US Food and Drug Administration agreed to let the company test the cells in patients who'd been recently injured. Asterias focused the current trial on patients with cervical injuries because these are closer to the brain, so new nerve cells have a shorter distance to grow to gain functional improvements. People with severe cervical spine injuries are typically paralysed below the level of the damage. The company's hope is to restore arm and hand function for people with such injuries, potentially making a tremendous difference to a person's independence and quality of life.

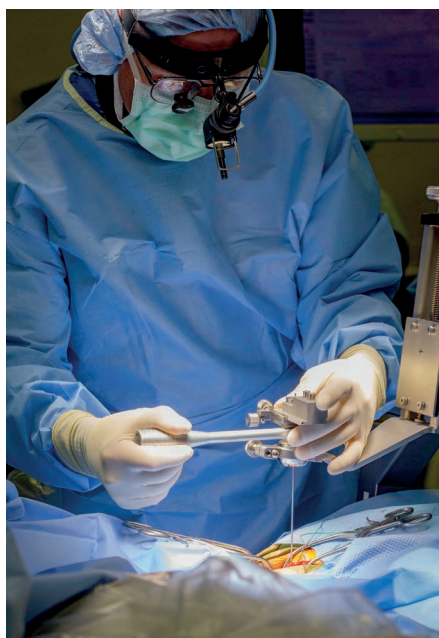
Asterias seems to have realized this hope in at least one patient who received one of the higher doses. Kristopher Boesen, who is 21, has had a dramatic recovery. In March, Boesen's car fishtailed in a rainstorm; he hit a telephone pole and broke his neck. About a month later, Boesen was still paralysed below the injury, and his neurological improvements seemed to have plateaued. His doctors at a trauma centre in Bakersfield, California, were in touch with Liu, who is an investigator in the Asterias trial. As soon as he was stable, Boesen travelled to Los Angeles to join the trial.

Liu injected Boesen's spinal cord with Asterias's cells in April. Two days later, Boesen started to move his hands, and in the summer, he regained the ability to move the toes on one foot.

Liu is excited about Boesen's response. "He was looking at being quadriplegic, and now he's

able to write, lift some weights with his hands, and use his phone," says Liu. "For somebody to improve like this is highly unusual — I want to be jumping out of my shoes." But Liu cautions that this is still a small trial, and that Boesen's response is just one anecdotal report. Until the results are borne out in a large, placebo-controlled clinical trial, Liu will remain earthbound.

The trial is currently recruiting between 5 and 8 patients for another cohort that will receive a doubled dose of 20 million cells. As the trial goes on, Asterias hopes to find clues about the underlying mechanism. "We're looking at changes in the anatomy of the injury," says the company's chief scientific officer, Jane Lebkowski. She says that there is some evidence that axons have traversed the injury site in patients who have recovered function. Preclinical work suggests that the cells might be sending growth-encouraging chemical signals to the native tissue. And, as support cells, the astrocytes may also be preventing more neurons from dying in the aftermath of the acute spinal injury.



A surgeon prepares to inject stem cells to treat a spinal injury as part of Asterias's clinical trial.

Not all clinical trials have performed so well. The SanBio and Asterias results are positive signals in a sea of negative or mixed trials. For example, StemCells of Newark, California, terminated its phase II trial of stem cells for the treatment of spinal-cord injury in May, and shortly afterwards announced that it will restructure its business. The company declined to comment for this article.

HOPE NOT HYPE

Physicians such as Liu and Steinberg temper their public enthusiasm about stem-cell therapies, so as not to give false hope to desperate patients. People with paralysing

injuries or those who have a neurodegenerative disease are easy marks for unscrupulous stem-cell clinics, whose therapies are not only unproven, but also come with risks.

"Patients say, 'Go ahead, doc, you can't make me any worse,'" says Keith Tansey, a neurologist and researcher at the Methodist Rehabilitation Center in Jackson, Mississippi, and president-elect of the American Spinal Injury Association. Unfortunately, that is not the case. Cell therapies given at a clinic, outside the context of a clinical trial, can lead to chronic pain, take away what little function a patient has left and render a patient ineligible for future studies, says Tansey. He has seen the consequences in his clinical practice. "I treated a kid who had two different tumours in his spinal cord from two different individuals' cells," he says.

Many unanswered questions remain about whether stem cells can heal the central nervous system in people, and how they might do it. Researchers also don't know what cells are the best to use. Is it enough for them to grow into supportive cells that send friendly growth signals, or is it better that they grow into replacement neurons? The answer is likely to differ depending on the site and nature of the disease or injury. If the stem cells are producing supportive factors that encourage growth and repair, it might be possible, says He, to discern what these are and give them directly to patients. But biologists are not yet close to deciphering the recipe for such a cocktail.

Tansey agrees that there are many unknowns — and these seem to be multiplying. "Every time we get an experiment done we realize it's more complex than we thought it would be," he says. Tansey thinks that the best way to resolve such uncertainties is with carefully regulated clinical trials. Rat models will only tell us so much — the human nervous system is much larger and is wired differently. If stem cells help patients such as Coontz and Boesen to regain their speech and give them greater independence without adverse effects, then it makes sense to continue, he says, even without knowing all the details of how they work.

Until these positive, but small, results are replicated in larger, controlled clinical trials, neurologists are containing their optimism. "I'd like to hear of any clinical trial that has more than an anecdotal benefit," says Tansey. And Liu is anticipating the day when he won't need to control his elation. In a few years, perhaps there will be a genuine opportunity to jump for joy. ■

Katherine Bourzac is a freelance journalist in San Francisco, California.

1. Lu, P. *et al.* *Cell* **150**, 1264–1273 (2012).
2. Lu, P. *et al.* *Neuron* **83**, 789–796 (2014).
3. Kadoya, K. *et al.* *Nature Med.* **22**, 479–487 (2016).
4. Steinberg, G. K. *et al.* *Stroke* **47**, 1817–1824 (2016).

PERSPECTIVE



Work with, not against, biology

Advances in technology have outpaced our understanding of organ development and injury response. It's time to reassess, says **Stephen Badylak**.

Medical science is failing to translate promising bench research on engineered tissues and organs into clinical practice. The original, and still dominant, template for creating replacement tissues was established in the 1980s and involves the culture of cells on a degradable scaffold in the shape of the target body part, which is then implanted into the recipient. Numerous reports of human ears¹, blood vessels², bladders³ and other body parts manufactured using this approach have generated a great deal of enthusiasm and, in the 1990s, gave rise to the field of tissue engineering.

But, with only a few exceptions, these replacement tissues are not being implanted in patients. There are the usual regulatory, manufacturing and reimbursement challenges, although most of these are manageable. One possible, and crucial, problem is that the tissues created by this generic approach lack many of the supportive elements that sustain the viability and function of the cells. Organs and tissues need effective vascular, lymphatic and innervation networks, but these are not part of the current design template; nor do they have enough time to develop organically before the transplant organ dies. Tissue engineers have also failed to consider the cells' microenvironment, which includes factors such as mechanical forces, pH and oxygen concentration, as well as signaling molecules embedded in the extracellular matrix. As a result, once implanted, the carefully engineered tissue is treated like a foreign body. Instead of being a drop-in, friendly replacement tissue, these grafts rapidly die from a lack of nutrients, appropriate growth factors and innervation⁴.

Then there is the immune system: the body's highly regulated method for orchestrating processes such as tissue and organ development, response to injury and infection, response to foreign cells and, importantly, regulation of tissue repair and reconstruction. Such processes may seem obvious to developmental biologists, immunologists and molecular biologists, but the tissue-engineering community has tended to either ignore the inevitable immune-system recognition of scaffolds and engineered tissues until the later stages of development, or conduct studies in immunocompromised rodent models that fail to take into account the potential effect of these factors on downstream clinical outcomes.

In spite of (or, in some cases, because of) incredible advances in genomic mapping, 3D cell printing, stem-cell biology, biomaterial development and bioreactor design, we have stubbornly adhered to the old approach. With current technology, a 3D replica of a patient's ear, based on a template created from elegant imaging techniques, can be fabricated by devices that deposit different types of cell layer by layer. As the 3D structure is recreated before our eyes, it is easy to forget that the natural intercellular connections necessary for cell-to-cell communication and support

are not present.

Failure to consider the effect of these robotic techniques on cell viability and phenotype, the omission of extracellular matrix molecules and the inability to 'print' the appropriate sensory and motor nerves and lymphatic networks, mean that these constructs will remain visually impressive but non-functional examples of state-of-the-art technology.

What can be done to advance the field and deliver on the promises of tissue engineering? It is time to combine forces with cross-disciplinary colleagues to first identify, then overcome, the biological obstacles that currently prevent the creation of functional tissue and organ replacements. These challenges are not insurmountable, but they must be acknowledged.

Of particular interest is the potential for tissue and organ regeneration that exists within the human genome (see page S58). Much has already been learned from salamanders, which can regrow an impressive array of body parts following injury⁵. The innate immune system has an active role in such developmental and regenerative processes⁶. Macrophages, for example, a cell type long identified with chronic inflammation, are necessary for limb regeneration in salamanders. This requirement suggests that it may be possible to reactivate similar signals in humans by working with the natural immune system rather than attempting to evade or suppress the host immune response. Successful tissue replacement may be as simple as placing an appropriate template embedded with biological cues at the intended anatomical site and using the body as the ultimate bioreactor, instead of attempting to recreate the organ or tissue externally.

Incredible progress has been made in the field of tissue engineering and regenerative medicine, but it is time for a course correction. The promises of the 1990s can still be fulfilled, but this will require willingness among tissue engineers to identify the processes that are working well, abandon the approaches that have inherent limitations and incorporate, where possible, the fundamental principles of normal mammalian development and regeneration. ■

Stephen Badylak is Professor of Surgery at the McGowan Institute for Regenerative Medicine in Pittsburgh, Pennsylvania, and deputy editor of *npj Regenerative Medicine*.
e-mail: badylaks@upmc.edu

1. Lee, S. J., Broda, C., Atala, A. & Yoo, J. J. *Biomacromolecules* **12**, 306–313 (2011).
2. Ren, X. et al. *Nature Biotechnol.* **33**, 1097–1102 (2015).
3. Horst, M. et al. *J. Tissue Eng. Regen. Med.* **7**, 515–522 (2013).
4. Simon, P. et al. *Eur. J. Cardiothorac. Surg.* **23**, 1002–1006 (2003).
5. King, B. L. & Yin, V. P. *PLoS One* **11**, e0157106 (2016).
6. Godwin, J. W., Pinto, A. R. & Rosenthal, N. A. *Proc. Natl Acad. Sci. USA* **110**, 9415–9420 (2013).

**INCREDIBLE
PROGRESS
HAS BEEN MADE
IN THE FIELD OF
REGENERATIVE
MEDICINE,
BUT IT IS TIME FOR A
COURSE CORRECTION.**



A tissue model is bioprinted using a 3D printer at Pennsylvania State University in University Park.

TECHNOLOGY

The promise of printing

Machines that deposit cell-laden inks are constructing tissues and organs, one layer at a time.

BY NEIL SAVAGE

People in their 30s are probably not thinking about needing surgery to repair their knees. But in a few decades, the rigours of daily living or the ravages of arthritis may send them to the operating theatre. And there, they might find a 3D printer ready to build new bone or cartilage in their bodies. Such is the hope of Ibrahim Ozbolat, a biological engineer at Pennsylvania State University in University Park who is developing 3D-printing techniques to repair tissue such as cartilage. He envisions a machine that will deposit successive layers of biocompatible material, laden with cells, into a defect. “In the future, we can have the patient under the bioprinter,” says Ozbolat. And his vision is not limited to knees. “Whatever section of the body needs to be fixed, the bioprinter can repair that.”

Many share Ozbolat's goal of producing living tissue in carefully designed shapes to repair or replace damaged parts of the body. As well as bone and cartilage, researchers are trying to develop skin, nervous-system tissue such as retinas, and even organs such as kidneys or a heart that would be printed externally and then transplanted into the body. But it may be a decade or more before even simple constructs make their way into clinical use. Before something as

complex as a kidney is possible, researchers will need to refine the materials, the implantation techniques and their ability to create complex structures such as vasculature. If 3D printing can be used to produce tissue on demand, it would revolutionize medicine, creating readily available transplant organs and tissues and eliminating the waiting list for replacement parts.

AN ENGINEERING MARVEL

In the engineering world, 3D printing is an umbrella term for a range of technologies that construct objects of almost any design by depositing layers of material. Working from a computer pattern, 3D printers can build complex objects that wouldn't be possible with more-conventional techniques, creating, for instance, internal contours where no tool could reach. They have become popular for generating quick prototypes, and are finding their way into the manufacturing processes of everything from medical device to aerospace industries.

The biological version is called 3D bioprinting, and it uses many of the same techniques, but adjusted to handle living cells. Often, the cells are contained within hydrogels — soft, jelly-like polymers that hold large amounts of water, but are viscous enough to hold their shape, at least for a while. The cells can also be suspended in a solution, like particles of pigment in printer ink.

The bioinks are sprayed through inkjet nozzles or extruded onto a base. They can be solidified by a change in temperature or pressure, or by adding chemicals or a wavelength of light. Once the right structure has been created, nutrients and growth factors are added to encourage the cells to develop into the right kind of tissue.

Researchers are testing combinations of techniques for printing a wide range of tissue types. Bone, for instance, grows best with a scaffold of cell-laden hydrogel, which orients the cells and gives them cues on where to grow. But cartilage, Ozbolat says, grows best when the cells determine their own arrangement, the way they do during embryonic development; a scaffold hinders the process. Cells within one hydrogel don't communicate well with those in another, and so don't provide the signals that cause them to grow together into a larger piece of tissue.

Ozbolat has developed a bioink based on alginate, a seaweed extract. He extruded thin strands of the ink containing cartilage cells, and these grew together, forming a large piece of tissue¹. He plans to use this technique with pancreatic cells to print islets of Langerhans, the insulin-producing parts of the pancreas, for transplantation into people with type 1 diabetes. And his technology works in living animals: by adding collagen to the bioink, he built bone and skin in the skull wounds of rats.

PATRICK MANSELL/PENN STATE UNIV.

Bioprinted cartilage could prove to be superior to both donor tissue and the plastic and titanium used in knee replacements, says Darryl D'Lima, director of orthopaedic research at the Scripps Clinic in San Diego, California. Although knee replacements made of plastic and metal last for 20 years, they're not as resilient as living tissue. "Metal and plastic and cement are strongest the day after you put them in," D'Lima says. "After that, they will only get weaker." And printed cartilage may turn out to be stronger than grafts from donated tissue. Not only can tissue be printed in the shape of the damage, avoiding the need to cut away healthy tissue, but the bioink should also fill the tiny crevices that no surgery could address, leading to greater overall strength. He admits that this is speculation; the technology isn't ready for comparative clinical trials. Researchers such as D'Lima are experimenting with 3D-printed parts in lab animals, but haven't yet proved that they're superior to conventional grafts.

D'Lima has been investigating 3D printing for the eyes, with the aim of treating the blindness caused by deterioration of the retina as it ages. He and Jeffrey Goldberg, an ophthalmologist at Stanford University in California, printed retinal ganglion cells onto scaffolding to see if they could grow retinas². Because the ganglion cells are neurons, they need to grow in a certain orientation. "We want their axons to point and grow in a particular direction because we want to get them to the right target," Goldberg says. The axons need to reach along the optic nerve to connect the retina to the brain. D'Lima and Goldberg managed this by creating a scaffold with a radial pattern that mimics the direction of the nerve fibres in the eyes, then printing the cells along those radial lines. Hydrogels that were rich in the fibrous protein laminin and contained some alginate helped to anchor the retinal cells in place, so that their natural signalling mechanisms could encourage them to grow in the right orientation. The researchers created structures with 72% of the axons oriented radially, compared with only 11% of those cultured in 2D on a plate.

One of the advantages of bioprinting for growing retinas is the specificity it affords. The retina contains two types of photoreceptor: rods and cones. Rods are more concentrated at the edge of the retina, whereas there are more cones in the centre. Bioprinting provides fine control over their placement. "You can't do that with any other technology," Goldberg says. It will probably be a few years before printed retinas can be tried as transplants, he says. One challenge is that nerve cells are packed more densely than, say, cartilage cells, and researchers are not yet sure how to achieve that density.

One line of research involves finding the best bioinks. Different materials have unique properties that affect whether they are easy to print with or hold their shape well, and whether they encourage cell growth. An ideal ink would stay

liquid, so it can be easily dispensed, but quickly turn into a more solid, gel-like structure without the use of chemicals or heavy doses of radiation that can harm the cells. "If you look at the technology of bioprinting, probably the limitation at the moment is the availability of bioinks," says Adam Perriman, a chemist at the University of Bristol, UK. Perriman has created a bioink that is a mixture of the hydrogel Pluronic (a mix of poloxamers) and alginate. This mix allows him to fine-tune the gelling time — Pluronic structures hold their shape well, but melt away too easily with changes in temperature, whereas alginate can gel too quickly. The mixed ink allows Perriman to print the structure he wants, then, once it's solidified, wash away the Pluronic, with the bonus that the gel leaves behind a network of micropores that allow the printed tissue to



Ear-shaped cartilage could be one of the first bioprinted tissues in clinical use.

take up nutrients³.

In another approach, Dong-Woo Cho, a mechanical engineer at Pohang University of Science and Technology in Gyeongbuk, South Korea, is using extracellular matrix to create bioink. He takes heart tissue, cartilage or fat and washes away all trace of cells, leaving behind just extracellular matrix, which consists of materials such as collagen and glycosaminoglycans⁴. He grinds up that material and mixes it with acetic acid and sodium hydroxide to make an ink that can be extruded in filaments, which gel when they're heated to body temperature. Cho thinks that, because the bioink is made from the matrix in which cells live, it will provide a more natural environment for the cells to thrive, and that it will be more biocompatible than inks made from materials not usually found in the body. He's using it to develop patches that could be helpful in the repair of heart damage.

Anthony Atala, director of the Wake Forest Institute for Regenerative Medicine in Winston-Salem, North Carolina, thinks that some simpler 3D-printed tissues could be in clinical use within the next few years. The first type to become available will probably be cartilage — it is a relatively flat structure that contains few cell types and does not require a blood supply. Atala has already printed cartilage and bone, and transplanted them into mice. In humans,

cartilage could be followed by hollow tubes such as an artery or urethra, and then hollow organs such as a bladder. "Of course, all tissues are complex, but the least complex are flat structures such as skin," Atala says.

A solid organ, however, may have a dozen or more cell types, and would need a vascular system to carry nutrients to its cells. Cells more than around 200 micrometres — a couple of human-hair widths — away from a source of nutrients quickly die, so creating a realistic vascular system will be necessary if researchers are to successfully build organs.

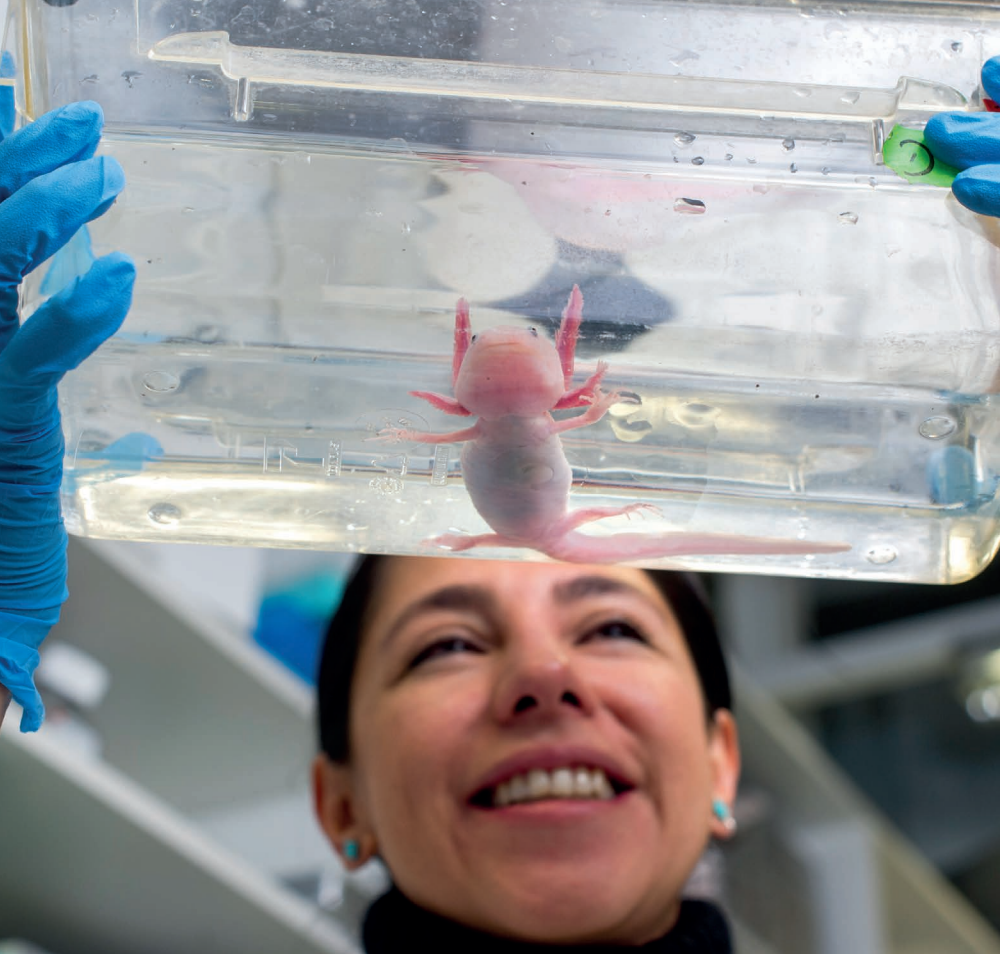
A team at Harvard University in Cambridge, Massachusetts, has taken the first steps to overcome that hurdle, printing thick tissue with a rudimentary vascular system and keeping it alive for weeks⁵. The team used three different inks: silicone to give a basic shape; a bioink infused with pluripotent stem cells that would turn into the tissue; and Pluronic, which is a gel at room temperature, but a liquid when cooled. The team printed the tissue, using the Pluronic to create threads throughout the tissue; after printing, the structure was cooled to 4°C, then the liquid was drained off to leave behind channels through which nutrients could flow.

Using this process, the Harvard team printed tissue one centimetre thick and kept it alive for more than six weeks. This was long enough for the stem cells to establish a bottom membrane of calcium phosphate, the base on which to grow bone. "There's really no limitation. We could go thicker," says Jennifer Lewis, the bio-engineer who led the research. The vasculature in this case was rudimentary; just a set of criss-crossing channels in successive layers. Actual organs, she says, will require a more complex pattern of veins and capillaries of varying sizes.

And bioprinting has the potential to do more than replace body parts with equivalent organs. Ozbolat is thinking about the prospect of enhancing the human body with new types of tissue. For instance, his lab has taken some preliminary steps towards creating an organ that converts chemical energy into electricity — a human version of what an electric eel can do. A person could have his or her own built-in, rechargeable battery to run anything from a pacemaker to a prosthetic limb. Even without printing body parts to give people superhuman abilities, bioprinting holds the promise of writing a new chapter in medicine. How many lives may be saved or bodies improved when it becomes possible to print an unlimited supply of replacement parts? ■

Neil Savage is a freelance science writer in Lowell, Massachusetts.

1. Yu, Y. *et al. Sci. Rep.* **6**, 28714 (2016).
2. Kador, K. E. *et al. Tissue Eng. A* **22**, 286–294 (2016).
3. Armstrong, J. P. K., Burke, M., Carter, B. M., Davis, S. A. & Perriman, A. W. *Adv. Healthc. Mater.* **5**, 1724–1730 (2016).
4. Pati, F. *et al. Nature Commun.* **5**, 3935 (2014).
5. Kolesky, D. B., Homana, K. A., Skylar-Scott, M. A. & Lewis, J. A. *Proc. Natl Acad. Sci. USA* **113**, 3179–3184 (2016).



ROBERT MICHAEL/AFP/GETTY

The Mexican salamander (or axolotl) can regrow limbs, making it an ideal model for studying regeneration.

ANIMAL MODELS

Unlock your inner salamander

Some animals can regrow body parts with ease. Biologists hope to figure out their secrets and apply them to humans.

BY SUJATA GUPTA

Picture this: a salamander and a mouse both lose a limb in an accident.

“A salamander laughs at that kind of injury,” says James Godwin, an immunologist and regeneration biologist at The Jackson Laboratory in Bar Harbor, Maine. “For a mouse, it’s devastating.”

That’s because, over the next few weeks, the salamander will regrow that lost limb — a perfect replica of the original — without any scar tissue forming. Meanwhile, the mouse, if it survives, will be left with a stump.

Researchers have long looked at these divergent trajectories and wondered whether it might be possible to make mice — and by extension humans — behave more like the salamander. This requires identifying the molecular and cellular processes activated during

the regenerative process that allow animals as varied as zebrafish, lizards and certain species of rodent to regrow everything from limbs to organs. The hope is to one day insert or reactivate the same pathways in humans.

It’s a tantalizing prospect. “Salamanders rebuild the tissue so perfectly that you can’t recognize that it was regenerated,” says Elly Tanaka, a regenerative biologist at the Research Institute of Molecular Pathology in Vienna.

Wouldn’t it be amazing, she says, if humans could do the same?

LOST ABILITY

The origin of animals’ ability to regenerate limbs and organs is uncertain. One possibility is that as mammals evolved more complex immune systems, they developed a scarring response to injury that helped to protect them against infection.

Essentially, says Kenro Kusumi, a regeneration biologist at Arizona State University in Tempe, an injured animal must ‘choose’ between a quick heal with scar tissue or a slow regenerative response that might leave it more vulnerable to predation for a few weeks. Although there are exceptions, mammals generally go for the quick fix, whereas salamanders and a few other vertebrates, such as fish and lizards, hunker down and rebuild.

If mammals have gained the scar response at the expense of regeneration, they might still have the molecular and cellular pathways needed for regeneration lying dormant. Regeneration in humans, says Kusumi, might “just be a matter of turning the switches back on.”

Humans, particularly early in life, do have limited repair capabilities, suggesting that some of the genetic programs for regeneration are present in mammals, but get turned off during development.

Consider the heart. When an adult has a heart attack, hundreds of millions of heart muscle cells die and a scar forms around the injury. With no way to regenerate or repair those cells, the weakened and scarred heart is more prone to subsequent heart attacks or failure. But research shows that newborn mice can regenerate heart-muscle cells. If you take a pair of scissors and snip off a third of the ventricle of the heart in a one-day-old mouse, says Godwin, the heart will repair itself without scarring. Do the same experiment at day seven, however, and that ability has been lost.

Intriguingly, human babies might have the same capacity. One case study reported an infant who had a heart attack at birth, and whose heart was scar-free a year later¹. Nevertheless, evidence is sparse and experiments are obviously unethical.

PUTTING PARTS TOGETHER

To achieve regeneration in humans, scientists must first identify the switches that turn on regeneration in other animals. One candidate for studying this is the salamander, which is capable of regrowing limbs that are perfect replicas of the original. That’s why Tanaka has been studying the amphibians for 20 years. “The salamander system provides a kind of guiding principle of how regeneration can work,” she says, “of how parts are put together.”

But it is not the only candidate. Over the years, regeneration biologists have expanded their animal repertoire, and their choice is often driven by the regenerating body part of interest. For instance, Kenneth Poss, a regeneration biologist at Duke University in Durham, North Carolina, works with zebrafish (*Danio rerio*) because he’s interested in heart regeneration. Although other animals, such as

salamanders and lizards, are also capable of rebuilding a heart, the main advantage of the zebrafish is that

NATURE.COM

To read more about the African spiny mouse: go.nature.com/2fa2kxt

it is an established lab animal, which makes it easier to create mutant and transgenic strains to test theories, Poss says.

Some researchers are turning their attention to lizards, which can generate new tails. Although lizards are much closer to humans, genetically speaking, than salamanders or zebrafish, the new tail is imperfect, composed entirely of cartilage and no bone. Even so, figuring out how to grow back cartilage in humans would be a huge feat. "Cartilage is not a tissue that most animals can even heal," says Thomas Lozito, a developmental biologist at the University of Pittsburgh in Pennsylvania, whose work compares regeneration in lizards and salamanders.

Thanks to efforts such as these, the basic steps involved in regeneration have been established in the organs of several types of animal; these include the zebrafish heart, salamander limb and mammalian liver. Although still a controversial idea, many scientists now think that regeneration is driven by a process known as dedifferentiation. That is, after an amputation or injury, adult fibroblast and muscle cells elsewhere in the body seem to receive signals to revert to a more embryonic state. "The cells go one step back in development," says Guo Huang, a regenerative biologist at the University of California, San Francisco.

As that dedifferentiation takes place, epidermal skin cells migrate to the injury site to seal it by forming a wound epidermis. Cells at the wound site then form a ridge, or apical epithelial cap, which sends out signals for the dedifferentiated cells to migrate in. Soon, the dedifferentiated cells start dividing to populate the bud of a new limb or other body part, known as a blastema.

Scientists have begun to identify and even manipulate the genes and proteins activated during regeneration. "We've identified enough molecules that we can kind of make a basic mechanistic description of what's happening in the salamander during regeneration," says Tanaka.

The next step, researchers say, is to put these genes back in lab rodents. Getting humans to regrow an amputated arm is still

a long way off, but researchers suspect that this intermediate step could happen within the next few years. "Once you translate a finding from a salamander or a fish to a mouse," Godwin says, "you've got a better chance of translating it to humans."

THE MISSING 99.8%

Perhaps that intermediate step in salamanders, lizards or fish will prove extraneous as researchers find mammals that are capable of the same sort of regeneration. Working in mammals would also circumvent the logistical problems that come with translating findings from distantly related species to humans.

Ken Muneoka, a regeneration biologist at Texas A&M University in College Station has been making this argument since the 1980s. Humans are very different from other creatures, Muneoka says. Whereas a salamander limb is mostly cartilage, a human limb contains a lot of bone. So any regenerative response in the mammal must first account for how to deal with big chunks of exposed bone. "That is very specific to the mammal," Muneoka says.

Some mammals do seem to have sorted out at least part of the secret to regeneration, however. Consider the story of the African spiny mouse.

In 2009, Ashley Seifert, a newly minted developmental biologist, spent the summer in Kenya. An acquaintance who was familiar with the country had told Seifert about a local rodent that could lose — and regrow — huge chunks of skin. Intrigued, Seifert set up some traps and captured several spiny mice (*Acomys*), which are abundant in the rocky outcroppings that pockmark the savannah. He quickly discovered that even just handling these rodents caused their skin to tear. Initially, the skin-healing process looked similar to that of other rodents, but after about 20 days, rather than the typical hairless scar, the spiny mice started to regrow hair follicles. In fact, the skin never scarred at all.

"When we conducted the initial experiments in the field, we watched the animals heal over time and regenerate the tissues," says Seifert, now at the University of Kentucky in Lexington. In a process known as autotomy, these rodents seem to be casting off body parts to evade a predator — the same sort of self-preservation method seen in lizards and salamanders. Seifert had found a mammal that seemed to have turned its regenerative switch for skin back on².

Seifert discovered that the mice could also repair holes in their ear — a classic body part in which to study wound healing — including regenerating skin, hair, muscle and even cartilage. His research (along with Muneoka's) suggests that the mice regenerate by first

developing a structure that bears a strong resemblance to the blastema seen in salamanders and other regenerating vertebrates. This supports the idea that regeneration in mammals such as the spiny mice evolved in the same manner as regeneration in salamanders.

Seifert's finding raised the enticing possibility that there may be other mammals that have abilities like the African spiny mouse. "There's around 5,300 mammal species on Earth right now," Seifert says. "Less than 0.2% have ever been surveyed for regenerative abilities."

GETTING TO HUMANS

Regenerating lost cartilage or appendages in humans is still decades or more away, but various lines of research suggest that turning those ancient regenerative switches back on in humans may come sooner.

Perhaps the most exciting clues for doing this comes from Muneoka's work in rats — specifically their digit tips. It's been known for at least half a century that when a child severs a finger above the top-most joint, that the fingertip (bone, flesh, nail and all) will typically grow back. This ability seems to become weaker with age, although teenagers and adults have also been known to grow back digit tips. Nobody knows for sure why this happens, but some suspect that the nail creates an environment that is conducive to regeneration or that mammals evolved the ability to regenerate in this region because the digits are crucial to limb use and survival. Whatever the reason, Muneoka wants to know how it happens.

Using mice, who display a similar phenomenon, Muneoka has shown that the gene *Msx1* controls a signalling protein known as Bmp. Remove either *Msx1* or Bmp and mice lose the ability to repair their digit tips. He has also found that adding Bmp stimulates partial regeneration of digits severed below the first joint³.

"We have shown that you can enhance the regenerative properties of mammals," Muneoka says.

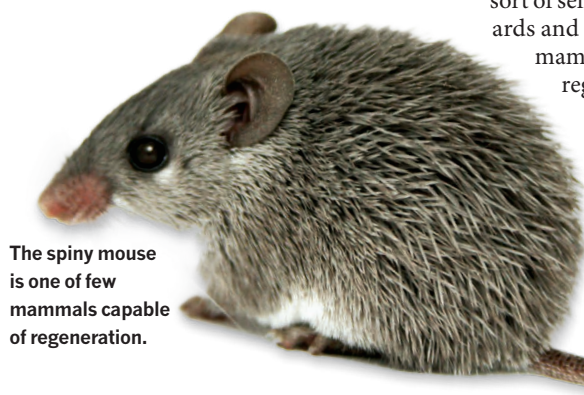
For Muneoka, that is proof of concept that regenerative capabilities can also be enhanced in humans. Initially, he suspects, the process will work better in children than in adults and will be limited to digits rather than entire limbs. "I believe that with adequate funding, human-finger regeneration in children will be possible within 20 years," he says. We may never catch up with the salamanders, but at least we're off the starting blocks. ■

Sujata Gupta is a freelance science writer in Burlington, Vermont.

1. Haubner, B. J. et al. *Circulation Res.* **118**, 216–221 (2016).
2. Seifert, A. W. et al. *Nature* **489**, 561–565 (2012).
3. Yu, L., Hana, M., Yana, M., Leea, J. & Muneoka, K. *Dev. Biol.* **372**, 263–273 (2012).

"Less than 0.2% of mammal species have ever been surveyed for regenerative abilities."

ASHLEY SEIFERT



The spiny mouse is one of few mammals capable of regeneration.



DIABETES

Encapsulating the problem

Cell therapy could cure type 1 diabetes — if only the immune system didn't get in the way.

BY ELIE DOLGIN

Douglas Melton allowed himself a brief moment of celebration. After 23 years of trying, he had finally managed to grow tissue in the laboratory that could replace the cell clusters, or islets, in the pancreas that are destroyed by type 1 diabetes (T1D) — the autoimmune disease that affects two of his children. Melton, co-director of the Harvard Stem Cell Institute in Cambridge, Massachusetts, had transformed embryonic stem cells into the specialized β -cells found in islets that sense glucose and secrete insulin to help to control blood sugar¹.

The ramifications were huge. The ability to produce a potentially limitless supply of β -cells meant that people with T1D would no longer need to receive islet transplants from deceased donors — an option that can help only a few hundred patients per year because of the dearth of donor organs. Most of the millions of adults and children living with T1D have to

monitor their glucose levels continually and inject themselves with insulin every day.

But after throwing a party for his lab group — at which Melton gave everyone blue winter hats emblazoned with his signature 'Skittle diagram', a schematic of colourful circles showing the developmental progression from stem cell to β -cell — reality set in. Even though he could now grow millions of β -cells in a single flask, an achievement for which he would be recognized with the 2016 Ogawa-Yamanaka Stem Cell Prize, Melton now faced a new challenge: how to protect implanted cells from the autoimmune attack that had destroyed the original cells. "Suddenly that problem loomed very large," says Melton, "because now it was the problem."

To prevent rejection of the islets — both because the cells come from unrelated donors and because of the recipients' islet-directed autoimmunity — patients have to take lifelong courses of immune-suppressing drugs. But these medicines come with serious

side effects, including increased risk of infection and cancer. Melton wanted to avoid these agents, which meant he needed some sort of device or material that would shield implanted β -cells — lab-grown or donor — from the immune system, while allowing nutrients such as glucose, insulin and oxygen into and out of these cells.

"The dream is to be able to build some kind of immuno-isolation device that will allow people to get the benefit of the cells without having to suppress their immune system," explains one of Melton's collaborators, Daniel Anderson, a bioengineer at the Massachusetts Institute of Technology (MIT) in Cambridge.

That dream has eluded academics, entrepreneurs and big pharmaceutical companies for more than 40 years. Thanks to the new-found ability to make β -cells from scratch, however, researchers now have a consistent and reliable cell source. This means that they

Spheres of sugar polymers designed to package engineered cells.

A. BADER/O. VEISEH/A. VEGAS/ANDERSON/LANGER LAB/KOCH INST. AT MIT

can methodically compare different encapsulation systems side by side, rather than tinkering through trial and error with whatever leftover β -cells they could get their hands on — the “dregs of material”, as Melton puts it. And this rational engineering approach is leading to improvements in the design of both large and small cell-safeguarding techniques.

Some companies, including the regenerative-medicine heavyweight ViaCyte, based in San Diego, California, are loading thousands of β -cells into macroencapsulation devices that are as big as the palm of your hand. Others, like the start-up Sigilon, based in Cambridge, Massachusetts, are parcelling up individual bundles of β -cells into microscopic shells smaller than a mustard seed. With both strategies, “we’re at this inflection point” where success is within reach, says Julia Greenstein, vice-president of discovery research at JDRF in New York City, a non-profit formerly known as the Juvenile Diabetes Research Foundation. “We’ve seen a much more scientifically directed approach to the problem than ever before.”

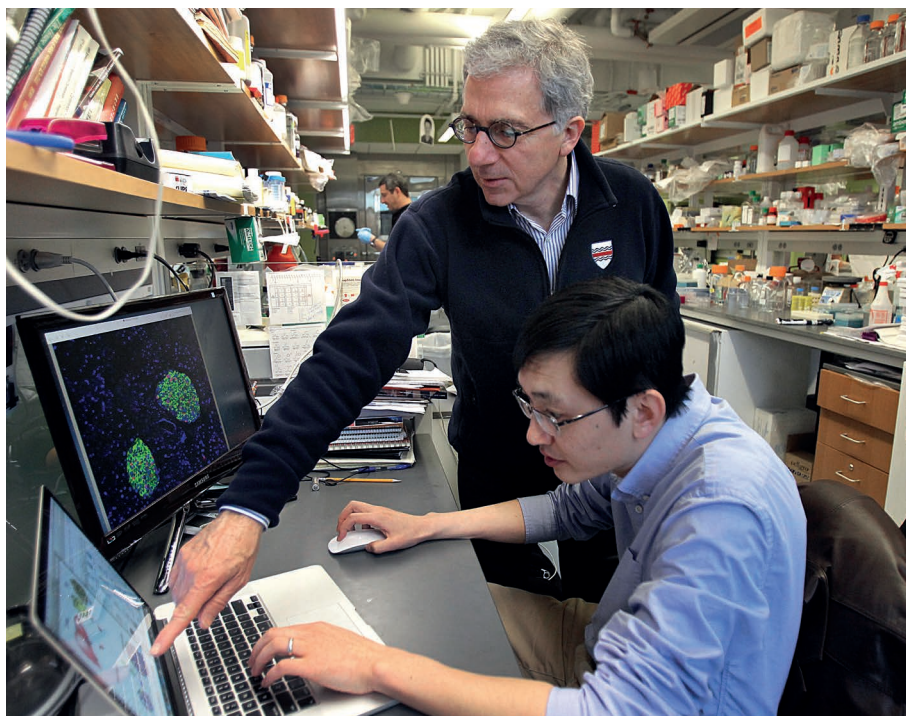
STEALTH MODE

For many years, the most popular capsule material for transplanted β -cells has been a seaweed extract called alginate. It’s been tested in rodents, dogs, monkeys and even humans in a few pilot clinical studies. The human trials showed that the material was safe, even in people who were not taking immunosuppressant drugs. But the therapeutic benefit was marginal because, within weeks of implantation, the alginate would usually begin to attract immune cells such as macrophages and neutrophils. This led to the deposition of fibrotic scar tissue that gummed up the capsules, choking off the cells inside.

This type of immune reaction was different from the one that originally destroyed the patients’ β -cells, but it was equally damaging to the prospects of this therapeutic approach. Seeking a derivative of alginate that could evade immune detection altogether, Anderson teamed up with his MIT colleague Robert Langer. The researchers systematically screened close to 800 chemical offshoots of alginate in mice. They found one variety — triazole-thiomorpholine dioxide alginate — that seemed to go completely unnoticed by the immune system².

Tiny spheres of this super-alginate survived for up to six months when implanted in macaques. And, when loaded with Melton’s stem-cell-derived β -cells, the capsules could restore typical levels of blood sugar in a mouse model of T1D, with no signs of an immune reaction³. “We demonstrated, for the first time, a material that remains fibrosis-free when implanted in the body,” says Omid Veisheh, a biomaterials researcher who worked on the project as a postdoc at MIT. “There hasn’t been anything like this.”

Veisheh will start his own lab in 2017 at Rice



Douglas Melton (back), co-director of the Harvard Stem Cell Institute, has found a way to engineer β -cells.

University in Houston, Texas. But until then, he is sticking around in Cambridge to help get Sigilon off the ground. Named after the Spanish word for stealth, Sigilon launched this year with US\$37.5 million in funding to commercialize the new alginate for a variety of biomedical applications. These include two potential ways to treat T1D. One is microencapsulated cell therapy. By 2018, Sigilon intends to show that this technology works in people using donor islets, and it is looking to partner with cell-therapy companies about then testing a stem-cell-derived β -cell enveloped in its alginate. The second approach is to use the material to coat parts of bionic pancreas systems that enter the body (see ‘Bionic versus beta’). “One way or another,” says Sigilon president and chief executive Paul Wotton, patients with T1D will “benefit from this platform.”

SMALL TALK

In the microencapsulation field, size matters. In tests with the super-alginate, the MIT team used capsules with a diameter of 1.5 millimetres, which it has demonstrated are much less immunogenic than the 0.5-mm capsules that researchers in the diabetes cell-therapy field have conventionally used⁴. But a tripling of the diameter means a nearly 30-fold increase in the volume of each capsule. And given the large number of capsules required to contain the hundreds of millions of β -cells needed to control a person’s diabetes, there are few places in the body where the therapy could be implanted.

The capsules probably won’t fit under the skin or in another easily retrievable location — and regulatory agencies have insisted, as a safety

measure, that any stem-cell-derived diabetes therapy implanted in patients should be fully recoverable. That’s why Alice Tomei, a bioengineer at the University of Miami in Florida, has developed what she calls a “shrink-wrapping technology”, which uses microfluidics to apply a thin biocompatible coating to clusters of cells to make the smallest possible capsule — one that’s only about 0.2 mm across⁵. Her material of choice, polyethylene glycol, may be more immune-reactive than Sigilon’s super-alginate, but Tomei argues that her thinly wrapped cells

“The dream is to be able to build some kind of immuno-isolation device.”

will be small enough to implant in more accessible spots in the body.

Tomei is evaluating her technology using Melton’s stem-cell-derived β -cells, in collaboration with start-up Semma Thera-

peutics in Cambridge, Massachusetts. Melton launched Semma in 2015 with the biotech investor and entrepreneur Robert Millman. (Millman’s wife came up with the name: a combination of Sam and Emma, the names of Melton’s two diabetic children.) Although Semma has its own encapsulation technology through the acquisition of drug-delivery-technology company Cytosolv, it is also looking for partners such as Tomei to test a range of encapsulation systems with the company’s stem cells. “Anyone who’s got a device, we’ll work with them,” says Millman, Semma’s chief executive, “because even with the best cells, if we don’t have the right device it’ll fail.”

Another of Semma’s collaborators is Beta-O₂ Technologies, a company based in

TECHNOLOGY

Bionic versus beta

In September 2016, the US Food and Drug Administration approved the first medical device that can both supply insulin when blood sugars are high and shut off insulin delivery when sugar levels start to drop, with little input from the patient. This artificial pancreas system — the MiniMed 670G from Medtronic, a medical device company headquartered in Dublin — combines two devices, one that monitors blood glucose levels and one that infuses insulin, with an algorithm that automates the delivery of insulin.

The bulk of the MiniMed set-up, including most of the electronics and the insulin pump, sits outside the body in a mobile-phone-sized device (pictured) that

people can carry in a pocket or wear on a belt. This device is attached to a tiny tube that is inserted beneath the skin. Much of the glucose-monitoring system is found in a patch worn on the surface of the abdomen, but this is also connected to a subcutaneous sensor.

The components that go into the body are prone to fibrotic immune reactions, and thus currently need to be replaced every few days. Coating the glucose sensor or the insulin-delivering catheter with a less immune-reactive material, such as the one being developed by Sigilon, a start-up in Cambridge, Massachusetts, could make the components last longer. But patients still need to tell the MiniMed (and similar technologies) when they've exercised or eaten — and no device can offer blood-glucose control that's anywhere near as tight as that provided by the body's own cells.

That's why Douglas Melton, co-director of the Harvard Stem Cell Institute in Cambridge, Massachusetts, who has two children with type 1 diabetes, continues to work on a cell-replacement therapy for the disease that would add back the insulin-producing β -cells that are lost. His benchmark for success: "It has got to make my children have a life in which they don't think any more about what they're going to eat or how much they've exercised any more than you and I do." **E.D.**

Rosh-Haayin, Israel, that has been working on encapsulated cell therapies for T1D for more than a decade. As the name suggests, Beta-O₂ was created to develop a way of delivering oxygen to implanted cells — a problem that's especially acute with larger macroencapsulation devices, which impose a larger barrier than capsule technologies between the blood supply that carries the oxygen and the energetically hungry β -cells.

The company's initial prototype was a chamber about the size of a hockey puck that is implanted under the skin and requires daily injections with oxygen. As proof of principle, Beta-O₂ tested this device by loading it with donor islets and implanting it in five patients in Germany⁶ and Sweden. The results were encouraging: the cells remained alive and healthy for months. Beta-O₂ is working on its second-generation device, which, according to chief executive Yuval Avni, will be able to hold more cells and require oxygen injections only

"Even with the best cells, if we don't have the right device it'll fail."

once a week. But the company needs a more reliable cell source, and Avni has high hopes for Melton's cells.

BOUNDING INTO THE CLINIC

Melton's original recipe for making β -cells was cumbersome. It took 35 days of carefully swapping 5 different growth media and mixing in 11 different factors, including sugars and proteins. According to Felicia Pagliuca, a former postdoc in Melton's lab who now leads cell-biology research and development at Semma, she and her team have dramatically streamlined the protocol. "We are leaps and bounds further from where we were," she says. And they have a strategy for getting the cells into clinical trials, even before an encapsulation device is ready.

The plan is to make β -cells from induced pluripotent stem cells created from people who need insulin, but whose bodies don't attack their own cells, which happens in people with T1D. Since there would be no tissue mismatch or chance of autoimmune reaction, those cells could then be implanted back into the patients without any immune-suppressing

drugs or barrier technologies.

Semma is focused on three patient populations, none of which have autoimmunity: people with a form of type 2 diabetes called lean diabetes, in which β -cells have stopped working; individuals who have had their pancreases surgically removed because of problems such as chronic inflammation; and patients with an insulin-dependent form of cystic fibrosis. The company hopes to test its cells in one of these populations in three to four years; trials with any sort of encapsulated device for people with T1D will follow at a later point.

But this means that Semma might be playing catch-up with its competitor ViaCyte. Earlier this year, the firm absorbed one of its chief rivals, a division of Johnson & Johnson called BetaLogics, while also announcing promising early data from the first human trial of an encapsulated stem-cell-derived product for T1D.

ViaCyte's PEC-EnCap device is made up of a semi-permeable pouch, about the size of a sticking plaster, that contains thousands of pancreatic precursor cells, each derived from embryonic stem cells. The company uses precursor cells, rather than fully mature β -cells, because these cells are hardier under the low-oxygen conditions that follow implantation, when the packets haven't yet integrated with the blood system. Over the past 2 years, ViaCyte has implanted its devices under the skin of about 20 patients without immunosuppression. In many recipients, the pancreatic precursors have grown into insulin-producing β -cells — although these cells often die after a few months, owing to a fibrotic immune reaction on the device exterior.

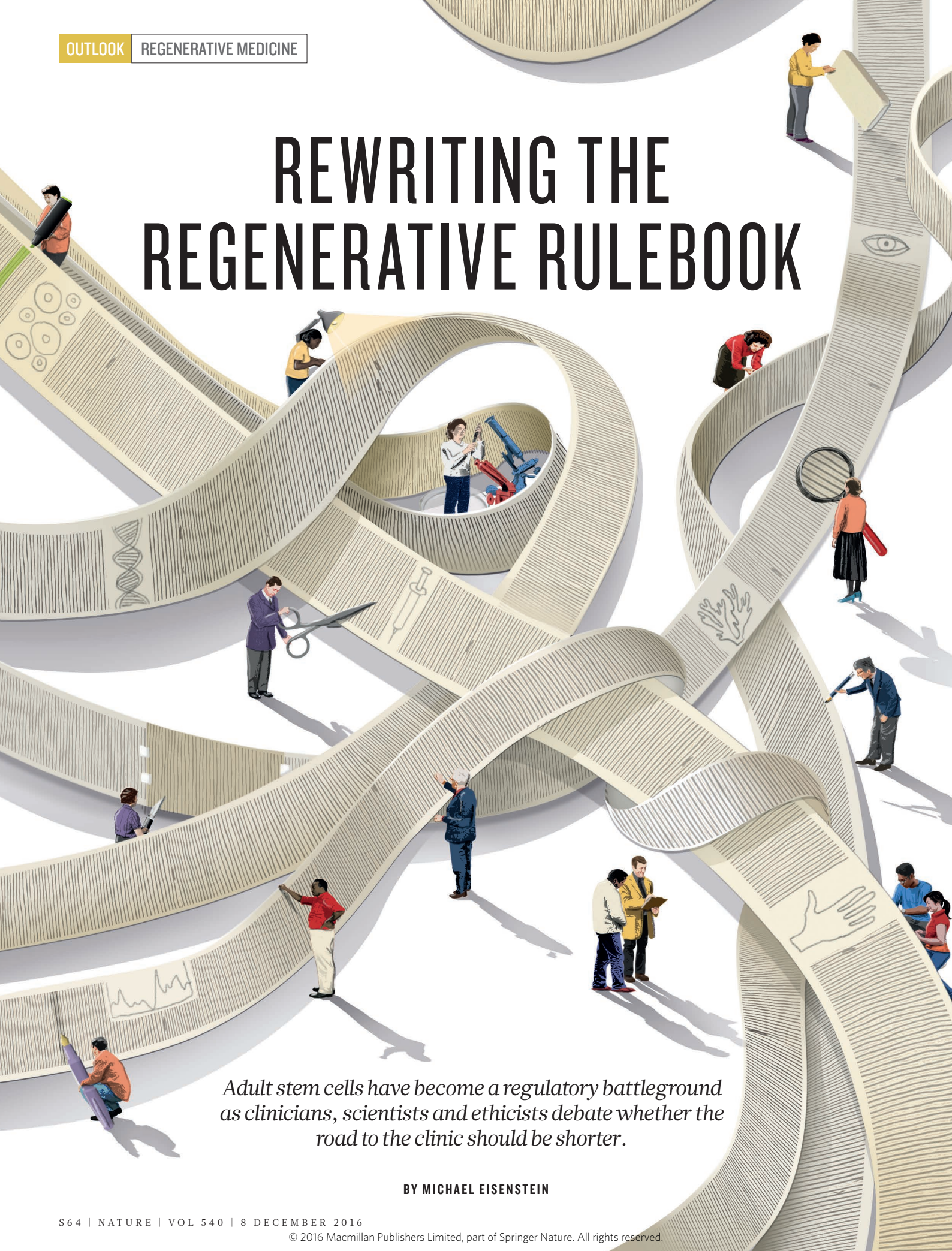
"That's a proof of feasibility that this is achievable, but we still have a lot of work to do," says ViaCyte's chief executive, Paul Laikind. "The goal now is to reduce or delay that foreign-body response long enough for the cells to engraft." ViaCyte hopes to achieve this by modifying either the encapsulation device or some other aspect of the treatment protocol before it moves into the next phase of testing with a full therapeutic doses of its product.

By then, perhaps, Melton's β -cells could also be ready for testing in patients with T1D. Melton is confident that with the right delivery system, these cells can cure his children's illness. "It just makes sense to me," Melton says, "that if you can make the cell that's missing in a person we ought to be able to find a way to put that cell back into people." ■

Elie Dolgin is a science writer in Somerville, Massachusetts.

1. Pagliuca, F. W. *et al.* *Cell* **159**, 428–439 (2014).
2. Vegas, A. J. *et al.* *Nature Biotechnol.* **34**, 345–352 (2016).
3. Vegas, A. J. *et al.* *Nature Med.* **22**, 306–311 (2016).
4. Veissh, O. *et al.* *Nature Mater.* **14**, 643–651 (2015).
5. Tomei, A. A. *et al.* *Proc. Natl Acad. Sci. USA* **111**, 10514–10519 (2014).
6. Ludwig, B. *et al.* *Proc. Natl. Acad. Sci. USA* **110**, 19054–19058 (2013).

REWRITING THE REGENERATIVE RULEBOOK



Adult stem cells have become a regulatory battleground as clinicians, scientists and ethicists debate whether the road to the clinic should be shorter.

BY MICHAEL EISENSTEIN

Shortly after Paul Knoepfler started his blog, e-mails began to flood his inbox. In 2010, his was one of few websites focused on making stem-cell science accessible to the public, and numerous people came looking for answers.

Knoepfler, a stem-cell biologist at the University of California, Davis, was well-versed in the scientific state of the field, but the inquiries he received took him aback. “They weren’t really asking about academic clinical trials,” he recalls. “They were almost all asking about stem-cell clinics.” He realized that people were desperate for information on the facilities that offered stem-cell preparations (which were often poorly defined) for an array of medical conditions, including many for which no medical benefits have been proved, such as autism spectrum disorder and Parkinson’s disease. At the time, these shadowy operations were based mainly in Asia, Latin America, Eastern Europe and the Caribbean, outside the reach of the US Food and Drug Administration (FDA) and the European Medicines Agency.

Today, the US stem-cell business is booming, and the FDA has a fight on its hands. In a recent study, Knoepfler and Leigh Turner, a bioethicist at the University of Minnesota in Minneapolis, counted 351 US-based businesses running 570 clinics that offer stem-cell treatments for cosmetic, orthopaedic, neurological and other conditions (L. Turner & P. Knoepfler *Cell Stem Cell* **19**, 154–157; 2016). The FDA has taken a hard line in its approach to stem-cell therapies, with regulations that generally call for the same rigorous testing procedures currently required for drugs and devices, but has done little to stave off the growth of the stem-cell-clinic industry. Meanwhile, some doctors and patients are pushing back, arguing that the agency is preventing medical progress with its overly strict and cautious approach to stem-cell regulation and impeding physicians’ right to practise medicine as they see fit.

The lines are now being drawn. Critics of the current regulations are challenging the fundamental aspects of stem-cell-therapy oversight and pursuing legislative solutions, whereas defenders of those rules favour a more conservative approach. “There’s this difficult-to-define ‘no man’s land’ between what is clearly the practice of medicine and what clearly isn’t,” says Charles Cox, a paediatric surgeon working in regenerative medicine at the University of Texas Health Science Center at Houston. “That’s where the tension arises.”

ADULT MATERIAL

The current dispute centres on a class of adult (as opposed to embryonic) stem cells that can be easily harvested and prepared for transplantation through a minor surgical procedure. These mesenchymal stem cells (MSCs) can be collected from a patient’s bone marrow or fat tissue, or from donor tissues, including the placenta and umbilical cord. MSCs have the capacity to form fat, bone or cartilage, making them potentially useful for repairing bones and joints. They also churn out biomolecules that can quell inflammation and stimulate repair in various tissues of the body. “It’s more than just bone regeneration — for example, we’re also spanning to neurological indications and autoimmune diseases,” says Massimo Dominici, who studies clinical applications of MSCs at the University of Modena and Reggio Emilia in Italy.

The accessibility and physiological properties of MSCs have made them the cell of choice for most clinics, but many of the ways in which these cells are used run afoul of FDA rules. Agency regulations establish

two broad categories for what it terms human cells, tissues and cellular and tissue-based products (HCT/Ps). Products that fall under a part of the rule known as section 361 are essentially treated like donor blood and organs. These regulations emphasize protecting recipients against transmission of infectious disease, but do not require extensive agency oversight.

Other HCT/Ps are governed by section 351. These are considered indistinguishable from drugs and must undergo a rigorous regulatory process before being given to patients. Some of the lines that separate section 351 products from those of section 361 are clearly drawn. For example, cells and tissues used homologously, meaning they perform the same function in the recipient as they do in the donor — such as the transplantation of bone marrow to restore healthy blood-cell production — are regulated under section 361. And therapies that employ a patient’s own stem cells (autologous) are more likely to fall under section

361 than those that use allogeneic cells from a donor. The situation becomes murkier in other areas. Any tissue that is more than ‘minimally manipulated’ is viewed as a drug. Ambiguities can arise, however, because separating stem cells from their neighbours always entails some degree of manipulation. “Sometimes you have to use enzymes to digest tissue and isolate the MSC precursors,” says Dominici. “In the past, this was considered minimal manipulation, but now the trend is to consider it a non-minimally manipulated procedure.”

The FDA’s authority over MSCs was tested in a high-profile lawsuit against Regenerative Sciences in Broomfield, Colorado. The treatment sold by the firm used autologous bone-marrow-derived MSCs to repair joint injuries. This is a plausible application for these cells, but the cells were being cultivated at a separate laboratory before re-implantation at the clinic. The FDA said that this represented more than minimal manipulation, and in 2014, the US Court of Appeals upheld the agency’s right to regulate the treatment. Many biologists agree with this interpretation of the regulations.

“Cell division in culture is nothing like when cells divide in the body,” says Lynn O’Donnell, a cell-therapy specialist at Ohio State University’s Wexner Medical Center in Columbus. “You’re accelerating the process, which can lead to increased risk for potential genetic alterations.”

RETICENT REGULATORS

Despite the court supporting the FDA’s interpretation of the regulations, the agency has not pursued legal action against any more stem-cell clinics. This has surprised many researchers who were expecting to see an increase in regulatory activity, says Turner. “That does not appear to have happened,” he says.

The FDA declined to comment for this article, but did issue a statement asserting its authority to respond to abuses of HCT/P regulations with “a variety of advisory, administrative and judicial actions depending on the particular violations”. These actions have generally taken the form of warning letters, which threaten further action if violations are not addressed. However, only a handful of letters have been issued to clinics, with no clear pattern. Turner notes that one recent letter targeted a clinic that offered only minor cosmetic procedures, while other businesses that make more ambitious medical claims continue to operate unimpeded.

Some doctors, however, think that they are being unfairly constrained. Ricardo Rodriguez is a cosmetic surgeon in Baltimore who heads the International Federation for Adipose Therapeutics and Science (IFATS), an organization that promotes research of therapies based on fat-derived

“PATIENTS
WERE MISLED
BY FALSE
INFORMATION
ABOUT THE
POTENTIAL OF
THESE CELLS.”

cells. He thinks that the FDA's position is grounded in excessive caution, motivated, in part, by past experiences in gene therapy — including when the death of a trial participant stalled clinical development for more than a decade. “The FDA made these arbitrary rulings many years ago that have very little to do with science and a lot to do with past experience and high-profile disasters,” says Rodriguez. He feels that the FDA regulations are overly broad and simplistic, denying physicians access to relatively safe therapeutic strategies by lumping them together with higher-risk regenerative-medicine approaches.

SETTING BOUNDARIES

During 2014 and 2015, the FDA issued four draft guidance documents detailing its interpretation of the HCT/P regulations, but these did not provide the level of clarity that many researchers and clinicians were hoping for. “I can wade through them for the most part, because I’ve been reading FDA stuff for years,” says Knoepfler. “But they’re not simple — they’re complicated, and not concrete.”

These documents have also stirred debate about fat-derived stem cells. This subpopulation of MSCs was identified in 2001, and are not referred to specifically in the original HCT/P regulations, which were issued the same year. However, these cells have since generated intense clinical interest — either in a purified form or as a heterogeneous, MSC-enriched preparation of fat known as the stromal vascular fraction. The FDA's draft guidance on adipose tissue brings these various formulations under a common umbrella, but critics argue that the agency's interpretation of this tissue is too narrow. “They are defining adipose tissue as solely structural — meaning that it cushions and supports, and that's all it does in the human body,” says Mary Ann Chirba, a legal scholar who specializes in health policy at Boston College Law School in Newton, Massachusetts. The draft guidance defines even a fat transplant to the breast for augmentation or postsurgical reconstruction purposes as non-homologous, on the premise that the biological purpose of the breast is milk production — something that is beyond the remit of adipose tissue. In Rodriguez's view, the FDA's definition of fat's physiological role “has no basis in biological reality”, and overlooks fat's broader endocrine, metabolic and other functions.

**“THE ONLY WAY
TO ADVANCE THIS
FIELD IS BY PASSING
LEGISLATION.”**

However, Knoepfler points out that most approaches for producing adipose MSCs — and particularly the stromal vascular fraction preparations commonly used in clinics — yield mixtures of cell types in addition to the desired stem cells that bear little resemblance to natural fat. These are, therefore, sufficiently manipulated to trigger stricter regulation regardless of how they are used. “I’m on the same page as the FDA in thinking that this is something that is not a natural product,” he says. However, Knoepfler also expresses some concern about the narrowness of the FDA's biological definitions. “I do not agree with their position on breast reconstruction,” he says.

Even with purified MSCs, questions remain about exactly what they are and how they work. Indeed, the biologist who coined the name mesenchymal stem cell in 1990 — Arnold Caplan, now at Case Western Reserve University in Cleveland, Ohio — has come to regret the term. “When I lecture, I beg the audience not to call them stem cells,” he says. Caplan points to evidence that MSCs are derived from precursor cells that surround blood vessels all over the body, where they stimulate local tissue repair when damage or disease occurs, rather than rebuilding the tissue themselves — the main feature associated with conventional stem cells. Because MSCs are active in myriad tissues and organs, Caplan views them through a different lens. “Based on what these cells do *in vivo*, I’d say you’ve got homologous use for MSCs in any tissue where you have a blood vessel,” he says.

Dominici's experience of using MSCs for bone regeneration has led him to a different conclusion. “Not all MSCs are capable of doing what we want them to do,” he says. MSCs derived from one tissue may exert different effects when transplanted into another part of the body, he says. Dominici thinks that a better way forward is to develop molecular assays to profile the therapeutic activity of different MSC preparations. This information could be more useful to clinicians in determining which cells are appropriate for a given disease than where the cells initially came from.

CONFLICT RESOLUTION

In September, the arguments about the FDA's oversight of stem-cell therapies were aired at a two-day hearing, at which speakers from almost every sector with a stake in stem cells discussed the draft guidelines. Chirba was among those advocating for a reconsideration of homologous-use rules. “Most FDA regulatory frameworks evaluate devices and drugs based on intended use and the intention of the person who is applying it,” she says. “I don’t understand why you can’t look at a cell or tissue the same way.” Many patients related personal tales of recovery and pleaded for easier access to the treatments that they saw as the source of their relief. Other speakers offered words of caution, including cell biologist Jeanne Loring of the Scripps Research Institute in La Jolla, California, who warned that “a lack of understanding or deliberate ignorance of regulations has led to increasing exploitation of desperate patients by incompetent clinics”.

The FDA may respond to issues raised at this hearing, but it is under no obligation to do so. Caplan believes that little will change. “The only way to advance this field is by passing legislation,” he says. Caplan helped formulate the REGROW Act, a bill to accelerate the entry of section-351-regulated cell and tissue therapies into the market. In its first iteration, the bill would have conferred conditional approval to HCT/Ps that demonstrated clinical safety and some data supporting efficacy — essentially eliminating phase III trials. Phase III testing is intended to provide robust proof of efficacy, but is also labour-intensive and expensive: one study estimates that costs range from US\$11.5 million to \$52.9 million (A. Sertkaya *et al. Clin. Trials* **13**, 117–126; 2016), depending on the field, and this is likely to be higher for a cutting-edge cellular therapy.



Stem cells are prepared before being given to a patient at a clinic in Mexico.

XINHUA/ALAMY



A protest in Rome in support of the controversial Stamina Foundation, which claimed that its stem cells could treat diseases such as Parkinson's.

Supporters of the bill think that it would level the playing field for companies and clinical researchers that lack the resources of big pharmaceutical companies. Japan adopted such a regulatory approach in 2014, and although too little time has passed to determine its pros and cons, Caplan says that “all our American companies have now set up shop in Japan”. However, Cox feels that the costs and bureaucratic hurdles have not thwarted his research team’s clinical efforts, and does not see these hurdles as an impediment to the growth of the field. “Some of it is a bit of a pain, but that hasn’t been a barrier,” he says. “It’s only a barrier if you don’t want to keep records and do things in a rigorous way.”

Japan requires post-market surveillance to prove safety and efficacy within a defined window of time, as a prelude to permanent approval, and this was an initial component of REGROW as well. However, the bill has divided patient advocates, with some applauding the idea while others remain wary of moving too quickly. “This could create a situation where patients and providers are paying for a product that’s not proven to be safe or effective,” says Bruce Bebo, executive vice-president for research at the National Multiple Sclerosis Society in New York City. “It lowers standards, and I think it would disincentivize more-rigorous research.” The bill was also opposed by the Alliance for Regenerative Medicine, a cell-therapy industry group in Washington DC, and the most recent version of it has abandoned conditional early approval to focus on more conventional strategies for accelerating clinical trials. The bill is still in committee, and may be further modified before coming up for a vote.

Both the European Union and the United States have compassionate-use exemptions that can give people with severe and untreatable diseases access to drugs in the clinical pipeline. Many stem-cell researchers support this loophole, but there are opportunities for abuse. For example, in 2013, Italy was rocked by a scandal surrounding a Brescia-based company called the Stamina Foundation, which

used a controversial trial to deliver unproven MSC-based treatments to people with incurable diseases such as Parkinson’s and muscular dystrophy. “Patients were misled by false information regarding the potential of these cells,” says Dominici, who presented evidence against Stamina to the Italian Senate.

Data suggest that MSC-based therapies are generally safe when applied under controlled clinical conditions. However, adverse events and even deaths have been reported after patients have visited stem-cell clinics, and these will inevitably become more regular as more people visit them. “One thing about medicine is that we should never underestimate our ability to make things worse,” says Cox. Indeed, one physician recently reported treating three women who became blind after receiving adipose-derived MSCs as a treatment for macular degeneration in a Florida-based clinic (H. Ledford *Nature* **537**, 148; 2016). But as Knoepfler discovered, convincing patients — particularly those with chronic or terminal disease — to wait for the science to advance, is a hard sell, and it has become clear that researchers and regulators need to do a better job of building bridges so that patients start perceiving regulatory agencies as allies rather than obstructions.

Research societies such as the International Society for Cellular Therapy in Vancouver, Canada, and the International Society for Stem Cell Research in Skokie, Illinois, are stepping up by providing educational materials to steer people away from clinics offering unproven treatments. However, without a concerted effort from regulators, this battle might be won by the slickest marketing — an area where commercial clinics have an edge. “A buyer beware mentality,” says Turner, “doesn’t do much to shield people from the most disturbing part of the marketplace.” ■

Michael Eisenstein is a freelance science writer in Philadelphia, Pennsylvania.